

**Diplomarbeit**

**Selbststudium in der Abdomensonographie mittels  
E-Learning im Vergleich zum Hands-on-Training**

eingereicht von

**Florian Tiefenböck**

zur Erlangung des akademischen Grades

**Doktor der gesamten Heilkunde**

**(Dr. med. univ.)**

an der

**Medizinischen Universität Graz**

ausgeführt am

**Institut für Medizinische Informatik, Statistik und  
Dokumentation**

unter der Anleitung von

**Univ.-Prof. Dr.med.univ. Josef Smolle**

Graz, am 01.08.2024

*Eidesstattliche Erklärung*

*Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.*

*Graz, am 01.08.2024*

*Florian Tiefenböck eh.*

## **Danksagungen**

An dieser Stelle möchte ich mich ganz herzlich bei denjenigen Menschen bedanken, die mir geholfen haben, diese Diplomarbeit fertigzustellen. Allen voran ist hier mein Betreuer Herr Univ. Prof. Dr. Josef Smolle zu nennen. Danke für die umfangreiche Betreuung, die zeitnahen Rückmeldungen und das hilfreiche Feedback.

Mein großer Dank gilt den Tutor:innen von Sono4you Graz und die vielen guten Freund:innen, die ich während meiner aktiven Zeit dort gewonnen habe. Die Dankesworte richten sich im Speziellen an Anita, Esta, Lisa und Georg, die Ersteller:innen des E-Learnings, welches ich freundlicherweise für meine Studie als Lehrmaterial verwenden durfte. Namentlich erwähnen möchte ich auch die Prüfer:innen Antonia, Frieda, Christian und Jakob sowie die beiden Schauspieler:innen Hanna und Viktoria, die alle direkt in diese Studie involviert waren.

Bedanken möchte ich mich auch bei der Medizinischen Universität Graz für die Verfügungstellung der Räumlichkeiten sowie die Ermöglichung der Dauerausleihe der mobilen Ultraschallgeräte.

Abschließend bleibt ein letztes Dankeschön an meine Eltern zu richten, die mir mit ihrer Unterstützung das Medizinstudium ermöglicht haben und immer an meiner Seite gestanden sind.

## Zusammenfassung

**Hintergrund:** Die Sonographie ist ein sicheres diagnostisches bildgebendes Verfahren, welches zunehmend von Ärzt:innen aus vielen Bereichen der Humanmedizin genutzt wird. Daraus bedingt sich eine steigende Nachfrage nach Ultraschallkursen auch bereits unter Studierenden der Humanmedizin. Diese Nachfrage können die Medizinischen Universitäten aus personellen und finanziellen Gründen derzeit nicht vollends abdecken. Auf der Suche nach alternativen Formaten von Ultraschallkursen, welche weniger personal- und kostenintensiv sind, wurde unter anderem auch das Selbststudium als Option erkundet.

**Methodik:** 39 Studierende nahmen an der randomisiert kontrollierten, beobachterverblindeten prospektiven Studie teil. Sie absolvierten drei Einheiten zu je zwei Stunden zur Abdomensonographie. Die Interventionsgruppe (n = 20) brachte sich dabei die Lehrinhalte überwiegend im Selbststudium bei, während die Kontrollgruppe (n = 19) einen traditionellen Kurs mit einem Tutor erhielt. Alle Proband:innen absolvierten eine Prüfung, bestehend aus zwei Prüfungsaufgaben, welche nach dem OSAUS-Bewertungsschema von je zwei unabhängigen, verblindeten Prüfer:innen beurteilt wurde.

**Ergebnisse:** Die Proband:innen der Interventionsgruppe erreichten im Mittel 27,6 ( $\pm$  6,3) Gesamtpunkte, in der Kontrollgruppe waren es durchschnittlich 30,3 ( $\pm$  6,2) Gesamtpunkte. Es zeigte sich keine signifikante Überlegenheit der Kontrollgruppe über die Interventionsgruppe, welche im Selbststudium gearbeitet hatte. Es gab keinen signifikanten Unterschied zwischen den beiden Geschlechtern. Der Studienfortschritt und die Gruppengröße korrelierte jeweils signifikant positiv mit den Gesamtpunkten, während die Korrelation zwischen Prüfungszeit und Gesamtpunkten signifikant negativ ausfiel.

**Diskussion:** Studierende der Humanmedizin können die Standardschnitte der Abdomensonographie im Selbststudium erlernen und waren in dieser Studie der Kontrollgruppe nicht signifikant unterlegen. Dies kann als Entscheidungsgrundlage dienen, Kapazitätsbeschränkungen in der Ultraschallausbildung, die auf personellen Engpässen beruhen, durch mehr Selbststudium zusätzlich zu traditionellen Kursen zumindest teilweise zu kompensieren.

## Abstract

**Background:** Sonography is a safe diagnostic imaging method that is increasingly being used by doctors from various medical disciplines. As a result, there is a growing demand for ultrasound courses, even among undergraduate medical students. Medical universities are currently unable to fully meet this demand due to staffing and financial constraints. In the search for alternative formats of ultrasound courses that are less labour and cost intensive, self-directed learning was explored as an option.

**Methods:** 39 medical students took part in the randomised controlled observer-blinded prospective study. They completed 3 units of 2 hours each on abdominal sonography. The intervention group (n = 20) mainly taught themselves through self-directed learning, while the control group (n = 19) attended a traditional course with a tutor. All subjects completed an assessment consisting of two examination tasks, which were assessed according to the OSAUS assessment scale by two independent blinded examiners.

**Results:** The subjects in the intervention group achieved an average of 27.6 ( $\pm$  6.3) points, while the control group achieved an average of 30.3 ( $\pm$  6.2) points. The control group did not perform significantly superior than the intervention group. There was no significant difference between the two genders. Study year of the individual student and group size each correlated positively with total points, while the correlation coefficient between examination time and total points was negative.

**Discussion:** Undergraduate medical students can learn the basics of abdominal sonography through self-directed learning and were not inferior to the control group in this study. These results can serve as a basis for deciding whether capacity restrictions in ultrasound training due to staff shortages should be at least partially compensated for by more self-directed learning in addition to traditional courses.

# Inhaltsverzeichnis

DANKSAGUNGEN .....	III
ZUSAMMENFASSUNG .....	IV
ABSTRACT .....	V
INHALTSVERZEICHNIS .....	VI
ABKÜRZUNGSVERZEICHNIS.....	IX
ABBILDUNGSVERZEICHNIS .....	X
TABELLENVERZEICHNIS .....	XI
<b>1. EINLEITUNG .....</b>	<b>1</b>
1.1 HINTERGRUND .....	1
1.2 DIDAKTIK IN DER SONOGRAPHIE.....	2
1.2.1 Lehrinhalte .....	2
1.2.2 Selbststudium.....	2
1.2.3 Mobile Ultraschallgeräte .....	4
1.2.4 Die Peer-Teaching-Initiative Sono4you Graz.....	5
1.3 BEWERTUNG PRAKTISCHER ULTRASCHALL-FÄHIGKEITEN .....	6
1.3.1 Relevanz .....	6
1.3.2 Das OSAUS-Bewertungsschema.....	7
1.3.3 Weitere kompetenzbasierte Bewertungsschemata .....	8
1.4 ZIELE DER ARBEIT .....	9
<b>2 METHODEN UND MATERIALIEN .....</b>	<b>10</b>
2.1 STUDIENPOPULATION .....	10
2.1.1 Einschlusskriterien.....	10
2.1.2 Ausschlusskriterien.....	10
2.1.2 Anmeldung und Zustimmung .....	10
2.1.3 Randomisierung .....	11
2.2 LEHRINHALTE .....	11
2.2.1 Erste Einheit .....	12
2.2.2 Zweite Einheit.....	13
2.2.3 Dritte Einheit .....	13
2.3 KURSABLAUF.....	14
2.3.1 Kontrollgruppe .....	14
2.3.2 Interventionsgruppe.....	15
2.4 MATERIALIEN.....	16

2.4.1 E-Learning .....	16
2.4.2 Räumlichkeiten.....	17
2.4.3 Technische Geräte.....	17
2.5 PRÜFUNG DES LERNERFOLGES .....	18
2.5.1 Prüfungsaufgaben.....	18
2.5.2 Prüfungszeit .....	19
2.5.3 Prüfungsablauf.....	20
2.5.4 Schallmodell .....	21
2.5.5 Dokumentation .....	21
2.5.6 Prüfer:innen .....	22
2.6 BEWERTUNGSSCHEMA .....	23
2.6.1 Kategorien.....	23
2.6.1.1 Bedienung des Ultraschallgeräts .....	24
2.6.1.2 Bildoptimierung.....	24
2.6.1.3 Systematik der Untersuchung .....	25
2.6.1.4 Interpretation der Bilder .....	25
2.6.2 Punktevergabe .....	26
2.6.3 Kalibrierungsvideos .....	26
2.6.3.1 Niveau der Videos .....	27
2.6.3.2 Ergebnisse .....	28
2.7 STATISTIK .....	29
2.8 ETHIK .....	30
<b>3 ERGEBNISSE .....</b>	<b>31</b>
3.1 FALLZAHL UND DROP-OUT-RATE.....	31
3.2 DEMOGRAPHISCHE DATEN.....	32
3.2.1 Geschlecht.....	32
3.2.2 Studienfortschritt .....	33
3.2.3 Gruppengröße.....	34
3.3 HAUPTZIELPARAMETER .....	36
3.3.1 Deskriptive Statistik.....	36
3.3.2 Unterschiede zwischen den Studiengruppen .....	37
3.3.3 Bestehensquote.....	37
3.4 GESCHLECHT .....	38
3.5 STUDIENFORTSCHRITT .....	40
3.6 GRUPPENGROÖE .....	42
3.7 PRÜFUNGSZEIT .....	43
3.7.1 Deskriptive Statistik.....	43
3.7.2 Unterschiede zwischen den Studiengruppen .....	44

3.7.3 Zusammenhang zu den Gesamtpunkten.....	44
3.8 PRÜFUNGSAUFGABEN .....	45
3.8.1 Verteilung der Prüfungsaufgaben innerhalb der Studiengruppen .....	45
3.8.2 Zeit in Abhängigkeit der Prüfungsaufgabe .....	47
3.9 PRÜFER:INNEN .....	48
3.10 KATEGORIEN .....	50
<b>4. DISKUSSION .....</b>	<b>52</b>
4.1 ERGEBNISSE .....	52
4.1.1 Fallzahl und Drop-out-Rate .....	52
4.1.2 Demographische Daten.....	52
4.1.2.1 Geschlecht .....	52
4.1.2.2 Studienfortschritt .....	53
4.1.2.3 Gruppengröße .....	54
4.1.2 Hauptergebnisse .....	54
4.1.3 Nebenergebnisse.....	55
4.1.3.1 Punkte in der Kategorie Gerät .....	55
4.1.3.2 Korrelation der Punkte in den Kategorien.....	55
4.1.3.3 Prüfungszeit.....	56
4.2 LIMITATIONEN.....	56
4.2.1 Methoden.....	56
4.2.1.1 E-Learning.....	56
4.2.1.2 Ultraschallgeräte .....	57
4.2.1.3 Interventionsgruppe.....	57
4.2.2 Prüfung des Lernerfolgs .....	58
4.2.2.1 Prüfer:innen .....	58
4.2.2.2 Prüfungsaufgaben .....	59
4.2.2.3 Kategorie Interpretation .....	60
4.2.2.4 Schallmodelle .....	60
4.2.2.5 Videodokumentation .....	61
4.2.2.6 Bestehensgrenze .....	61
4.2.3 Übertragbarkeit der Ergebnisse .....	61
4.3 AUSBLICK .....	62
4.4 SCHLUSSFOLGERUNG .....	63
<b>LITERATURVERZEICHNIS .....</b>	<b>64</b>

## **Abkürzungsverzeichnis**

B-QUIET	Brightness Mode Quality Ultrasound Imaging Examination Technique
EFSUMB	European Federation of Societies for Ultrasound in Medicine and Biology
FAST	Focused Assessment with Sonography for Trauma
FOAMed	Free Open Access Medical Education
DEGUM	Deutsche Gesellschaft für Ultraschall in der Medizin
DOPS	Direct Observation of Procedural Skills
GRS	General Rating Scale
SDL	Self-Directed Learning
ÖGUM	Österreichischen Gesellschaft für Ultraschall in der Medizin
OSAUS	Objective Structured Assessment of Ultrasound Skills
OSCE	Objective Structured Clinical Examination
TSC	Tasc Specific Checklist
WFUMB	World Federation of Societies for Ultrasound in Medicine and Biology

# Abbildungsverzeichnis

<i>Abbildung 1: Flussdiagramm der Lehrinhalte .....</i>	<i>12</i>
<i>Abbildung 2: Flussdiagramm des didaktischen 4-Stufen-Schemas .....</i>	<i>15</i>
<i>Abbildung 3: Beispielbild aus einer Prüfungssituation (mit freundlicher Genehmigung der Schauspieler:innen) .....</i>	<i>22</i>
<i>Abbildung 4: Flussdiagramm der Proband:innenanzahl .....</i>	<i>32</i>
<i>Abbildung 5: Häufigkeit der Geschlechter in den Studiengruppen .....</i>	<i>33</i>
<i>Abbildung 6: Studienfortschritt in den Studiengruppen .....</i>	<i>34</i>
<i>Abbildung 7: Gruppengrößen in den Studiengruppen .....</i>	<i>35</i>
<i>Abbildung 8: Boxplot der Gesamtpunkte (A) bzw. der Punkte in den Kategorien (B) in Abhängigkeit der Studiengruppe .....</i>	<i>36</i>
<i>Abbildung 9: Bestanden und nicht bestanden Proband:innen in Abhängigkeit der Studiengruppe .....</i>	<i>38</i>
<i>Abbildung 10: Boxplot der Gesamtpunkte in Abhängigkeit des Geschlechts .....</i>	<i>39</i>
<i>Abbildung 11: Streudiagramm der Gesamtpunkte in Abhängigkeit des Studienfortschritts (in Jahren) .....</i>	<i>41</i>
<i>Abbildung 12: Streudiagramm der Gesamtpunkte in Abhängigkeit der Gruppengröße .....</i>	<i>42</i>
<i>Abbildung 13: Streudiagramm der Gesamtpunkte in Abhängigkeit der Prüfungszeit (in Minuten) .....</i>	<i>45</i>
<i>Abbildung 14: Boxplot der Zeit pro Aufgabe (in Minuten) in Abhängigkeit der Prüfungsaufgabe .....</i>	<i>48</i>
<i>Abbildung 15: Streudiagramm der Gesamtpunkte bei den beiden Prüfer:innen .....</i>	<i>50</i>
<i>Abbildung 16: Streudiagramme der Punkte zwischen den beiden Prüfer:innen in den Kategorien (A) Gerät, (B) Bild, (C) System, (D) Interpretation .....</i>	<i>50</i>

## **Tabellenverzeichnis**

<i>Tabelle 1: Kurzbezeichnung und Wortlaut der Prüfungsaufgaben</i> .....	19
<i>Tabelle 2: Bewertung der Kalibrierungsvideos</i> .....	29
<i>Tabelle 3: Drop-out nach der Randomisierung in Abhängigkeit der Studiengruppe</i> .....	31
<i>Tabelle 4: Häufigkeiten der Geschlechter in Abhängigkeit der Studiengruppe</i> .....	33
<i>Tabelle 5: Studienfortschritt (in Jahren) der Proband:innen in Abhängigkeit der Studiengruppe</i> .....	34
<i>Tabelle 6: Anzahl der Proband:innen in Kleingruppen mit verschiedener Gruppengröße</i> 35	
<i>Tabelle 7: Punkte in Abhängigkeit der Studiengruppe</i> .....	36
<i>Tabelle 8: Ergebnisse des Mann-Whitney-U-Tests der Unterschiede der Punkte zwischen den Studiengruppen</i> .....	37
<i>Tabelle 9: : Bestehensquote in Abhängigkeit der Studiengruppe</i> .....	38
<i>Tabelle 10: Punkte in Abhängigkeit des Geschlechts sowie Ergebnisse des T-Tests für unabhängige Stichproben</i> .....	39
<i>Tabelle 11: Gesamtpunkte in Abhängigkeit des Studienfortschritts</i> .....	40
<i>Tabelle 12: Ergebnisse der Rangkorrelationsanalyse nach Spearman zwischen dem Studienfortschritt und den Punkten</i> .....	41
<i>Tabelle 14: Gesamtpunkte in Abhängigkeit der Gruppengröße</i> .....	42
<i>Tabelle 15: Ergebnisse der Rangkorrelationsanalyse nach Spearman zwischen der Gruppengröße und den Punkten</i> .....	43
<i>Tabelle 16: Prüfungszeit (in Minuten) in Abhängigkeit der Studiengruppe</i> .....	44
<i>Tabelle 17: Korrelation nach Spearman zwischen der Prüfungszeit (in Minuten) und den Punkten</i> .....	45
<i>Tabelle 18: Häufigkeiten der Prüfungsaufgaben</i> .....	46
<i>Tabelle 19: Zeit pro Aufgabe (in Minuten) in Abhängigkeit der Prüfungsaufgabe (M = Mittelwert, SD = Standardabweichung)</i> .....	48
<i>Tabelle 20: Ergebnisse der Rangkorrelationsanalyse nach Spearman zwischen den Punkten der beiden Prüfer:innen je Proband:in</i> .....	49
<i>Tabelle 21: Ergebnisse der Rangkorrelationsanalyse nach Spearman zwischen den Punkten der jeweiligen Kategorien in der Studienpopulation</i> .....	51

# 1. Einleitung

## 1.1 Hintergrund

Neben Fachärzt:innen für Radiologie nutzen auch Ärzt:innen anderer Fachrichtungen zunehmend die Sonographie als wichtige diagnostische Maßnahme. Durch die Weiterentwicklung mobiler Ultraschallgeräte bezüglich Qualität, Preis und Benutzerfreundlichkeit ergänzt die Sonographie den physikalischen Status in immer größerem Ausmaß. Von einigen Autor:innen wird deshalb der Begriff „Sonoscope“ oder „Echoscope“ propagiert, um auf die Relevanz der Sonographie in der Erstliniendiagnostik auf einem Niveau mit dem Stethoskop hinzuweisen [1].

Ultraschall ist eine sichere diagnostische Bildgebung ohne Verwendung ionisierender Strahlung. Ein Schallkopf erzeugt durch Anlegen einer elektrischen Spannung an piezoelektrischen Kristallen hochfrequente Schallwellen im Ultraschallbereich. Diese Schallwellen werden im Körper an Grenzflächen von Geweben mit verschiedenen akustischen Impedanzen reflektiert. Die Informationen aus den reflektierten Echos werden vom Ultraschallgerät in Graustufenbilder umgewandelt [2].

Da Verfügbarkeit, Qualität und Anwendungsgebiete von Ultraschallgeräten stetig steigen, steigt auch die Nachfrage nach entsprechenden Ausbildungsangeboten. Die EFSUMB (European Federation of Societies for Ultrasound in Medicine and Biology), die Dachorganisation der europäischen Fachgesellschaften für diagnostischen Ultraschall, empfiehlt deshalb die frühestmögliche Implementierung der Ultraschalllehre in universitäre Curricula [1]. Jedoch mangelt es dafür häufig an der ausreichenden Verfügbarkeit sowohl fachlich als auch didaktisch ausgebildeten Personals, welches meist stark in den klinischen Alltag und die medizinische Forschung involviert ist [3]. In einer Umfrage unter europäischen medizinischen Universitäten wurde 2020 die mangelnde Zeit, die fehlende Finanzierung sowie fehlendes universitäres Personal als Hauptkriterien für den langsamen Fortschritt bei der Implementierung curricularer Ultraschallkurse erhoben [4].

## **1.2 Didaktik in der Sonographie**

### **1.2.1 Lehrinhalte**

Im „Klinischen Lernzielkatalog Österreichs“ formulierten die öffentlichen medizinischen Universitäten Österreichs in einem Delphi-Konsens ein national einheitliches Ausbildungsprofil für Medizinstudierende. Dabei wurde im Lernziel „RA047“ festgelegt, dass Absolvent:innen des Diplomstudiums Humanmedizin die Standardschnitte der Abdomensonographie auf Lerntiefe 2 beherrschen sollte. Diese Lerntiefe entspricht der selbstständigen routinierten Anwendung in der klinischen Praxis ohne Anleitung [5].

Trotz dieser gemeinsamen Absichtserklärungen und des ungebrochen steigenden Interesses an Ultraschallkursen fanden Prosch et al. (2020) in einer Umfrage unter 46 Universitäten in 17 europäischen Ländern, dass nur 56% der Universitäten praktische Ultraschallfertigkeiten lehrten und diese nur in 15% der Fälle diese Fähigkeiten praktisch abgeprüft wurden [4].

Laut einem Positionspapier der WFUMB (World Federation of Societies for Ultrasound in Medicine and Biology), der weltweiten Dachorganisation der Ultraschall-Fachgesellschaften, sollte ein Curriculum zur Ultraschalllehre mit der Vermittlung grundlegender physikalischer Prinzipien, Gerätekunde, Bildoptimierung und Sicherheitsaspekten starten. Anatomie, Physiologie, Untersuchungstechnik und häufige Pathologien sollten ebenfalls im Laufe des Kurses behandelt werden [6]. Die DEGUM (Deutsche Gesellschaft für Ultraschall in der Medizin) veröffentlichte zudem Richtlinien, welche Standardschnitte bei einer Abdomensonographie durchgeführt werden sollen [2].

Tarique et al. (2017) beschrieben in einem Review, dass Kursstruktur und Inhalt von Ultraschallkursen für Medizinstudierende nicht standardisiert sind und zwischen den Medizinischen Universitäten deutlich differieren [7]. Die EFSUMB hielt wiederum fest, dass aufgrund der uneinheitlichen Curricula der medizinischen Universitäten Europas ein einheitliches Ultraschall-Kursformat mit standardisierten Inhalten nicht realisierbar sei. Stattdessen sollten die medizinischen Fakultäten eigene Kurse nach den Empfehlungen der EFSUMB entwickeln, welche sich den lokalen Gegebenheiten anpassen [1].

### **1.2.2 Selbststudium**

Knowles (1975) formulierte eine übersichtliche und häufig referenzierte Definition des Selbststudiums, welches im englischsprachigen Raum als „Self-directed Learning“ (SDL) bezeichnet wird. Es handle sich um einen Prozess, in dem die Lernenden individuell die

Initiative ergreifen, um Lerndefizite zu eruieren und Ziele zu formulieren. Auch das Identifizieren von Ressourcen sowie das Suchen nach geeigneten Strategien, um die Interventionen umzusetzen soll selbstständig erfolgen. Abschließend evaluieren die Lernenden den Lernprozess in einer Selbstbeurteilung. Die Lehrenden sollen dabei nach Knowles lediglich moderierend und nicht wissensvermittelnd auftreten [8]. Es wurde gezeigt, dass das Selbststudium die intrinsische Lernmotivation der Studierenden fördert und so dafür sorgt, dass Erlerntes länger behalten wird [9]. Die drei essentiellen Merkmale, welche ein „Self-directed Learning“-Konzept haben sollte, sind nach Knowles:

- Auftreten der Lehrperson als Moderator:in, anstatt als Wissensvermittler:in
- Involvierung der Lernenden in die Auswahl der Ressourcen und Strategien
- Selbstbeurteilung des Lernerfolgs [8].

In der Praxis der didaktischen Forschung auf dem Gebiet der Sonographie wird eine große Bandbreite an Konzepten als „Self-directed Learning“ bezeichnet, auch wenn sie nicht alle der oben genannten Kriterien erfüllen [10]. Murad et al. (2010) zeigten in einer Metastudie, dass das Selbststudium in den Gesundheitswissenschaften effektiver war, wenn die Lernenden in die Wahl der Lehrmaterialien involviert waren. Außerdem profitierten fortgeschrittene Lernende mehr vom Selbststudium als Anfänger:innen. Die Interaktivität der Lernmaterialien hatte jedoch keinen signifikanten Effekt auf den Lernerfolg [10].

In der Ultraschalldidaktik ist das Konzept des Selbststudiums noch ein vergleichsweise neuer Ansatz. Es ist die traditionelle Annahme, dass zum Erlernen der praktischen Fähigkeiten im Ultraschall Hands-on-Training unter Supervision von Tutor:innen nötig sei [11]. Es existieren einige wenige Arbeiten zum Selbststudium in der Sonographie. Diese haben häufig explorativen Charakter ohne Vergleich zu einer Kontrollgruppe [12,13]. Auch die Proband:innenanzahl ist meist sehr überschaubar und wurde nicht mittels einer a priori Poweranalyse ermittelt [13,14]. In einzelnen Studien wurde auch das Selbststudium an Ultraschall-Simulatoren untersucht [14–16]. Klare Aussagen zu optimalen Lernszenarien können im Bereich der Ultraschalldidaktik im Allgemeinen selten getroffen werden, da zu wenig methodisch hochwertige Studien existieren.

Das Selbststudium muss überdies vom Tele-Learning abgegrenzt werden. Letzteres wurde in der Zeit der COVID-19-Pandemie verstärkt als Alternative zum persönlichen Kontakt

genutzt. Es beruht zur Wissensvermittlung jedoch ebenso wie traditionelle Kurse auf dem Einsatz von Tutor:innen [17,18].

Fuchs et al. (2018) zeigten, dass Studierende, welche die fokussierte Echokardiographie im Selbststudium mittels eines E-Learnings mit portablen Ultraschallsonden erlernten, gleich gut abschnitten, wie Studierende, die einen traditionellen Kurs erhielten. Die Autoren selbst empfahlen in ihrer Conclusio jedoch die Implementierung des Selbststudiums lediglich als Zusatz zu den bereits bestehenden traditionellen Kursstrukturen [19].

Bei Signor et al. (2024) wurden je ein mobiles Ultraschallgerät pro Person ausgeteilt, welches die Studierenden in einer achtwöchigen Famulatur an einer Abteilung für Innere Medizin verwenden konnten. Während der Famulatur mussten Bilder von vorgegebenen Schnittebenen zur Beurteilung eingeschickt werden. Die Studierenden konnten ihre Punkte in der Bildqualität im Verlauf der Studie deutlich steigern, ein Vergleich zu einer Kontrollgruppe wurde jedoch nicht durchgeführt [12].

Bei Canty et al. (2018) schnitten jene Proband:innen, welche die fokussierte Echokardiographie im Selbststudium an Simulatoren erlernten besser ab, als jene, die einen eineinhalbtägigen Workshop erhielten [16].

Cawthorn et al. (2014) verglichen drei didaktische Konzepte, um die fokussierte Echokardiographie an Simulatoren zu erlernen, miteinander. Hier erzielten die Proband:innen nach dem Selbststudium signifikant weniger Punkte bezüglich der Bildqualität als jene aus einem traditionellen Kursformat oder einem Flipped-Classroom-Format [15].

### **1.2.3 Mobile Ultraschallgeräte**

Mobile Ultraschallgeräte werden von Seiten der EFSUMB als Geräte mit einer Bildschirmgröße eines Smartphones oder eines Tablets, welche einfach transportierbar sind, definiert [20]. Sie wurden in einigen der Referenzarbeiten zur Ultraschalldidaktik bereits als Ersatz zu traditionellen Ultraschallgeräten eingesetzt [12,17–19].

Bezüglich der Verwendung in der Lehre sieht die EFSUMB in der Verwendung mobiler Ultraschallgeräte das Potential, patientennahen Unterricht zu forcieren und durch die geringeren Kosten das Angebot an Ultraschallkursen erweitern zu können. Es wurde die Empfehlung ausgesprochen, mobile Ultraschallgeräte erst nach dem Erlernen der Grundlagen an traditionellen Geräten zu verwenden [20].

An der Medizinischen Universität Graz gibt es für Studierende der Humanmedizin seit 2022 das Angebot, an der Universitätsbibliothek mobile Ultraschallsonden für bis zu 72 Stunden auszuleihen. Diese können dann im privaten Rahmen oder auch in klinischen Praktika genutzt werden, um bereits erworbenes Wissen zu festigen oder sich neue Fertigkeiten anzueignen.

#### **1.2.4 Die Peer-Teaching-Initiative Sono4you Graz**

Sono4you ist eine studentische Initiative zur Förderung der Ultraschalldidaktik. Sie wurde 2007 in Wien von Studierenden als Antwort auf die steigende Nachfrage nach Ultraschallkursen bei mangelndem curricularem Angebot gegründet. Seit 2013 ist sie an der Medizinischen Universität Graz aktiv. In Kleingruppen zu je fünf Teilnehmer:innen, zwei Tutor:innen und zwei traditionellen Ultraschallgeräten werden die grundlegenden praktischen Fähigkeiten der Sonographie gelehrt. Dies passiert im Flipped-Classroom-Format, bei dem sich die Studierenden vorab mittels eines E-Learnings theoretisch vorbereiten. In den drei Einheiten zu je zwei Stunden steht dann das praktische Üben im Vordergrund. Es werden Grundkurse in Abdomen- und Notfallsonographie sowie Aufbaukurse in Echokardiographie und Kopf-Hals-Sonographie angeboten. Alle Kurse werden von den Tutor:innen ehrenamtlich abgehalten und von den Teilnehmenden in der Freizeit absolviert. Sie können von allen Studierenden der Medizinischen Universität Graz, die den ersten Studienabschnitt erfolgreich abgeschlossen haben, kostenlos in Anspruch genommen werden und sind als Ergänzung der curricularen Ultraschalllehre zu sehen. Aufgrund der räumlichen und personellen Limitationen kann auch das Angebot der Anfrage nicht zur Gänze gerecht werden und so werden die Kursplätze verlost [21].

Celebi et al. (2010) zeigten, dass studentische Tutor:innen effektiv in der Lehre der Abdomensonographie eingesetzt werden können [22]. Weitere Studien kamen zu ähnlichen Resultaten [11]. Auch die WFUMB empfiehlt den Einsatz studentischer Tutor:innen in der Ultraschalllehre [3]. Da die studentischen Tutor:innen jeweils wieder eigene Kolleg:innen ausbilden können ist dies ein zusätzlicher Ansatz, um der personelle Knappheit an didaktisch geschulten Lehrpersonen entgegenzutreten [11]. Weitere Vorteile eines solchen Peer-Teaching-Konzeptes sind die lockere Lernatmosphäre unter einer flachen Hierarchie, welche die Barrieren für Fragen senkt und so für höheres Engagement unter den Teilnehmenden sorgt [11]. Durch die geringen sozialen und intellektuellen Unterschiede zwischen Teilnehmenden und Lehrenden kann Peer-Teaching in manchen Bereichen sogar dem Unterricht durch universitäres Personal überlegen sein [3].

## 1.3 Bewertung praktischer Ultraschall-Fähigkeiten

### 1.3.1 Relevanz

Die Qualität einer Ultraschalluntersuchung und der dokumentierten Bilder ist stark von der Kompetenz des Untersuchenden abhängig. Aus diesem Grund ist es notwendig, feststellen zu können, wann eine Person ausreichend Kompetenz auf einem Gebiet aufweist [23].

Es wurden verschiedene Bewertungsschemata zur Beurteilung praktischer Ultraschall-Fähigkeiten publiziert. Höhne et al. (2022) boten in einem Review einen Überblick über die gebräuchlichsten Schemata. Am häufigsten wurden Checklisten und Skalen benutzt. Bei „Task-specific Checklists“ (TSC) werden Items auf einer Liste binär nach deren Erfüllung benotet und zu einer Gesamtpunktezahl addiert. Bei „General Rating Scales“ (GRS) werden Punkte in verschiedenen Kategorien meist auf einer Likert-Skala vergeben. Daneben finden sich auch noch die „Objective Structured Clinical Examination“ (OSCE), bei der im Stationsbetrieb klinische Situationen simuliert werden. Die „Direct observation of procedural Skills“ (DOPS) ist eine Evaluierung durch direkte Supervision realer Untersuchungen echter Patient:innen im Arbeitsumfeld. Die Selbstbeurteilung mittels Fragebögen ist eine Möglichkeit, die Motivation der Studierenden noch weiter zu steigern, liefert jedoch lediglich subjektive Daten [24]. Sie ist ein Hauptkriterium der engeren Definition von Selbststudium nach Knowles (1975) [8]. Häufig werden auch schriftliche Tests, meist im Multiple-Choice-Format, als Surrogatparameter benutzt. Dabei wird jedoch keine praktische Fähigkeit direkt bewertet [24].

In einem systematischen Review von Studien, welche bei medizinischen Simulationen sowohl eine TSC als auch eine GRS verwendeten, fanden Ilgen et al. (2015), dass die Skalen eine durchschnittlich höhere Inter-item-Reliability als die Checklisten hatten. Außerdem war mit Skalen eine feinere Differenzierung der Kompetenzlevel möglich. Während eine TSC eher prüft, ob eine Aufgabe vollständig erledigt wurde, ist die Qualität der Ausführung mit einer GRS besser zu differenzieren. Eine GRS erfordert jedoch Prüfer:innen mit ausreichend Expertise und Schulung auf das konkrete Schema, um den Gesamteindruck adäquat beurteilen zu können [25]. Unter Umständen wird bei einer TSC die Vollständigkeit der Durchführung höher bewertet als die Qualität der Ausführung. Durch die Entwicklung von Checklisten mit speziellen kritischen Items kann verhindert werden, dass viele Punkte mit eher trivialen Items gesammelt werden, während ein ernster Fehler in einer wichtigeren Kategorie sich wenig auf die Gesamtpunkte auswirkt [26].

Häufig wird die Kompetenz von Ärzt:innen lediglich anhand der Fallzahl bestimmter Untersuchungen bemessen. So sind beispielsweise zum Erreichen der Stufe 1 Zertifizierung der ÖGUM (Österreichischen Gesellschaft für Ultraschall in der Medizin) im Bereich der Abdomensonographie neben der Absolvierung der geforderten Kurse 500 unter Supervision durchgeführte Abdomenultraschall-Untersuchungen nötig, von denen mindestens 25% Pathologien aufweisen mussten [27]. Studien zeigten jedoch, dass Lernende einen individuell unterschiedlichen Lernbedarf haben und verschiedene Teilbereiche eine verschieden steile Lernkurve aufweisen [23]. Aus diesem Grund ist die kompetenzorientierte Prüfung praktischer Fähigkeiten einer willkürlich festgesetzten Anzahl an Untersuchungen vorzuziehen [23].

Es lässt sich aus der vorhandenen Literatur keine eindeutige Empfehlung für eine spezifische Art der Überprüfung oder ein spezifisches Bewertungsschema ableiten. Es ist anzunehmen, dass aufgrund der diversen, nicht einheitlichen Curricula darauf angepasste Prüfungen für jede medizinische Fakultät extra konzipiert werden müssen. Eine Kombination aus verschiedenen Bewertungsschemata ist eine Möglichkeit, die verschiedenen Limitationen zu kompensieren [24]. Im Folgenden wird eine Auswahl verschiedener validierter, kompetenzorientierter Bewertungsschemata vorgestellt.

### **1.3.2 Das OSAUS-Bewertungsschema**

Tolsgaard et al. (2013) entwickelten ein Bewertungsschema namens „Objective Structured Assessment of Ultrasound Skills“ (OSAUS), welches unabhängig von der konkreten Aufgabenstellung in jedem Fachgebiet zur Bewertung praktischer Ultraschallfähigkeiten verwendet werden kann. Die Untersuchung verschiedener Organe kann so geprüft werden, ohne das Schema spezifisch anpassen zu müssen. Dazu wurde ein Konsens unter 60 Expert:innen unterschiedlicher Fachrichtungen nach der modifizierten Delphi-Methode gefunden. In dieser „General Rating Scale“ (GRS) können in sieben Kategorien je bis zu fünf Punkte auf einer Likert-Skala vergeben werden [23]. Die konkrete Implementierung in dieser Studie wird im Abschnitt 2.6 erläutert.

Die Zuverlässigkeit des OSAUS-Schemas wurde von Todsén et al. (2015) bei einer Analyse von 24 Proband:innen, welche vier chirurgische Patient:innen untersuchten, validiert. Dabei konnte das Bewertungsschema sicher zwischen Anfänger:innen, Fortgeschrittenen und Expert:innen differenzieren. Außerdem korrelierten die Gesamtpunkte signifikant mit der

Anzahl der korrekten Diagnosen. Die Autor:innen empfehlen deshalb dieses Bewertungsschema zur Zertifizierung von Ärzt:innen vor der eigenständigen Tätigkeit [28].

Das OSAUS-Schema ist in der didaktischen Forschung im Bereich der Sonographie das am besten untersuchte Bewertungsschema und wurde von vielen der Referenzarbeiten verwendet [17,18,28–31]. Im Folgenden werden einige weitere alternative Assessment-Methoden vorgestellt, welche allerdings weniger verbreitet und weniger gut untersucht sind.

### **1.3.3 Weitere kompetenzbasierte Bewertungsschemata**

An der Heinrich-Heine-Universität Düsseldorf wurde über viele Jahre ein spezielles Kursangebot zur Abdomensonographie entwickelt, welches mit einer OSCE abgeschlossen wird. Hier wird neben praktischen Stationen auch das Zeichnen von schematischen sonographischen Schnittbildern aus dem Gedächtnis gefordert [32]. Aufgrund der sehr spezifischen Abstimmung der Prüfung auf die konkreten Kursinhalte ist eine Reproduzierbarkeit in einem anderen Setting wohl nur eingeschränkt möglich.

Bahner et al. (2011) entwickelten das Bewertungsschema „Brightness Mode Quality Ultrasound Imaging Examination Technique“ (B-QUIET), bei der ausschließlich die dokumentierten Bilder beurteilt werden [33]. Es ist daher ergebnisorientierter als die anderen vorgestellten Schemata, welche den Entstehungsprozess und die Einbindung in ein klinisches Setting mit in Betracht ziehen. In zwei Studien wurde dementsprechend auch keine signifikante Korrelation zwischen den Punkten aus B-QUIET und OSAUS gefunden [17,18].

Das Bewertungsschema „Quality of Ultrasound Imaging and Competence Score“ (QUICK-Score) ist eine Kombination aus einer Checkliste und einer Skala zur Beurteilung der Fähigkeiten in der Notfallsonographie. Im Speziellen soll damit das FAST-Schema („Focused Assessment with Sonography for Trauma“) bewertet werden. Das Bewertungsschema konnte in einer Pilotstudie sicher Expert:innen von Anfänger:innen unterscheiden [34].

Black et al. (2016) konzipierten in einem modifizierten Delphi-Konsens eine Skala, welche spezifisch auf die Beurteilung der praktischen Ultraschallfertigkeiten von Medizinstudierenden ausgelegt ist. In neun Kategorien können je ein bis fünf Punkte vergeben werden. Zusätzlich wurden drei Checklisten für definierte Organsysteme entwickelt [35].

## **1.4 Ziele der Arbeit**

Alternative Kursformate mit geringerem personellem Aufwand auf Seiten der Lehrenden werden aufgrund der steigenden Nachfrage nach sonographischen Lehrangeboten relevanter. Die Hypothese lautet, dass dazu das Selbststudium mit frei ausleihbaren mobilen Ultraschallgeräten und kostenlos zugänglichen Lehrmaterialien einen wichtigen Beitrag leisten kann.

Diese Arbeit evaluiert in einer Pilotstudie den Lernerfolg von Proband:innen nach einem Selbststudium in der Abdomensonographie und vergleicht diesen mit einer Kontrollgruppe, welche einen traditionellen Kurs mit einem Tutor absolviert. Beide Studiengruppen verwenden mobile Ultraschallgeräte und ein frei zugängliches E-Learning als Lehrmaterialien. Der Lernerfolg wird mit dem validierten Bewertungsschema OSAUS erhoben.

Es soll damit festgestellt werden, ob ein höheres Ausmaß an Selbststudium beim Erlernen der Abdomensonographie möglich ist, ohne dass es zu einer Verschlechterung des Lernerfolgs kommt. Sollte dies gelingen, wäre dies eine Entscheidungsgrundlage dafür, Kapazitätsbeschränkungen in der Ultraschallausbildung, die auf personellen Engpässen beruhen, durch mehr Selbststudium zumindest teilweise zu kompensieren.

## **2 Methoden und Materialien**

### **2.1 Studienpopulation**

#### **2.1.1 Einschlusskriterien**

Als Grundpopulation wurden alle Studierenden des Diplomstudiums Humanmedizin an der Medizinischen Universität Graz definiert. Voraussetzung war die erfolgreiche Absolvierung des ersten Studienabschnittes, welche die ersten beiden Studienjahre umfasst. So verfügten alle Proband:innen über das physiologische und anatomische Vorwissen, welches für die Erlernung der Abdomensonographie unerlässlich war.

#### **2.1.2 Ausschlusskriterien**

All jene Studierende, die bereits einen weiterführenden extracurricularen Kurs zur Abdomensonographie besucht hatten, waren von der Anmeldung für die Studie ausgeschlossen. Dies galt insbesondere für Kurse, welche bei Sono4you Graz absolviert wurden.

Weiters sollte keine tiefgehende praktische Erfahrung in der Abdomensonographie im Rahmen von Famulaturen oder anderen Praktika erfolgt sein. Die Grenze hierfür wurde bei maximal fünf selbstständig abdomensonographisch untersuchten Patient:innen festgesetzt.

#### **2.1.2 Anmeldung und Zustimmung**

Sämtliche relevanten Informationen zur Studie wurden über eine eigens eingerichtete Website veröffentlicht [36]. Über diese war auch die Anmeldung mittels eines Online-Formulars neun Tage lang möglich. Interessierte Studierende, welche die Einschlusskriterien erfüllten, konnten ein oder zwei Wunschtermine aus einer Liste von 12 Startterminen auswählen. Zum Anmeldezeitpunkt waren die Starttermine noch keiner Studiengruppe zugeteilt. Die Proband:innen wurden etwa zwei Wochen vor Beginn der ersten Kurseinheit über ihren fixierten Termin und das Ergebnis der Randomisierung informiert.

Am Ende des Anmeldeformulars gaben alle Proband:innen durch das Setzen eines Häkchens im entsprechenden Feld des Online-Formulars ihre Einwilligung zu der Teilnahme an der Studie bekannt. Die Interessent:innen wurden darüber aufgeklärt, dass ihre erhobenen persönlichen Daten (Vorname, Nachname, Geschlecht, Studienjahr, E-Mail-Adresse und

Telefonnummer) zur Kontaktaufnahme passwortgesichert gespeichert wurden. Die statistische Datenverarbeitung erfolgte pseudonymisiert mittels zufällig zugeteilten Identifikationsnummern. Außerdem stimmten sie zu, dass sie sich selbst als Schallmodell der Kleingruppe zur Verfügung stellen würden und dass während der Prüfung ihre Handbewegungen und das Ultraschallbild audiovisuell aufgezeichnet werden würden. Diese Einwilligung konnte zu jedem Zeitpunkt ohne Angabe von Gründen widerrufen werden.

Es meldeten sich 50 Personen an, wovon zwei Studierende ausgeschlossen wurden. Eine Person machte unvollständige Angaben, eine weitere Person hatte bereits einen extracurricularen Abdomensonographie-Kurs bei Sono4you Graz absolviert. Somit nahmen 48 Proband:innen an der Randomisierung teil (96%).

### **2.1.3 Randomisierung**

Die Interessent:innen wurden entsprechend ihren Wunschterminen in 10 Kleingruppen zu drei bis sechs Studierenden eingeteilt. Die oberste Priorität war dabei, allen Personen einen Termin anbieten zu können. Die Konstanz der Gruppengröße war zweitrangig, es wurde jedoch darauf geachtet, dass die Gruppengröße sechs Personen nicht überschritt und drei Personen nicht unterschritt, um ein Verhältnis von zwei bis drei Proband:innen pro Ultraschallgerät zu wahren.

Um die oben genannten Prioritäten zu erfüllen, erfolgte die Randomisierung ganzer Kleingruppen und nicht individueller Proband:innen. Es erfolgte keine Stratifizierung nach Geschlecht, Studienfortschritt oder anderen Merkmalen.. Ein nachträgliches Tauschen der Kleingruppe bei kurzfristiger Verhinderung war innerhalb derselben Studiengruppe möglich.

## **2.2 Lehrinhalte**

Die Lehrinhalte dieser Studie richteten sich nach den Inhalten des E-Learnings, welches in Abschnitt 2.4.1 näher erläutert wird. Dieses umfasst neben den physikalischen Grundlagen der Abdomensonographie ausgewählte Standardschnittebenen nach den Richtlinien der DEGUM (Deutsche Gesellschaft für Ultraschall in der Medizin) [37]. Abbildung 1 zeigt eine Übersicht über die vermittelten Lerninhalte, welche im Folgenden näher beschrieben werden.

Die Darmsonographie, welche von den DEGUM-Richtlinien als Lehrinhalt empfohlen wird, war nicht Teil des E-Learning und wurde deshalb in dieser Studie nicht gelehrt. Auf die

Darstellung des Pankreas wurde verzichtet, obwohl es im E-Learning behandelt wird. Die Darstellung des Pankreas im modifizierten Oberbauchquerschnitt ist bei nicht-nüchternen Schallmodellen erfahrungsgemäß deutlich eingeschränkt und mit hohen interindividuellen Unterschieden behaftet.

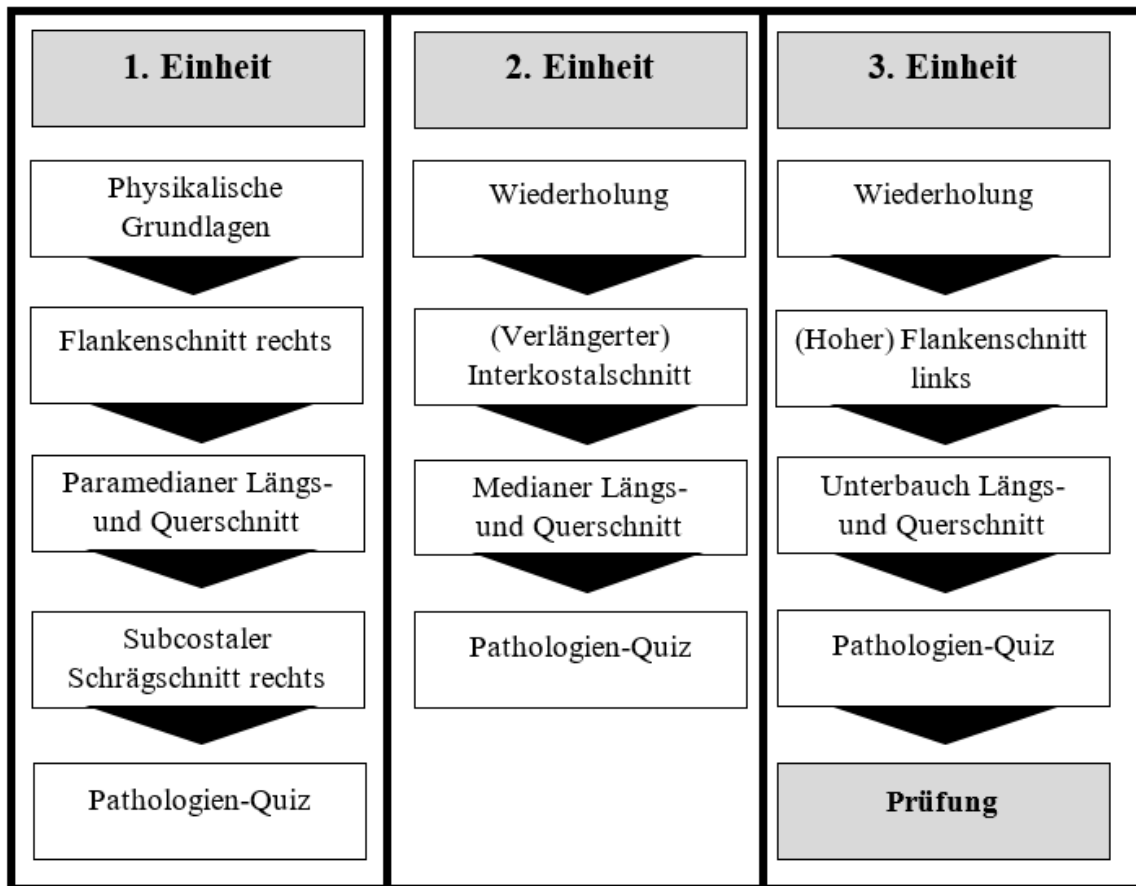


Abbildung 1: Flussdiagramm der Lehrinhalte

### 2.2.1 Erste Einheit

Die erste Einheit startete mit einer Einführung in den Hintergrund und die Fragestellung der Studie sowie mit einer Vorstellungs- und Kennenlernrunde, um eine lockere Lernatmosphäre zu schaffen. In diesem ersten organisatorischen Block erfolgte auch die individuelle Vereinbarung der beiden Folgetermine mit der Kleingruppe. Danach wurden die physikalischen Grundlagen des Ultraschalls und die Entstehung diverser Bildartefakte behandelt. Ebenso waren die Gerätekunde und die Schallkopfführung sowie die Orientierung am Bild Teil des einführenden Blocks.

Nach dieser etwa 15-minütigen Einführung startete der praktische Teil des Kurses. Zu Beginn wurde die Darstellung der rechten Niere sowie des Recessus hepatorenalis im Längs- und Querschnitt an der rechten Flanke geübt. Danach folgte die Darstellung der Gallenblase im paramedianen Längs- und Querschnitt rechts. Abschließend stand die Durchmusterung der Leber im subcostalen Schrägschnitt rechts mit Fokus auf der Darstellung der intrahepatischen Gallengänge und Lebervenen auf dem Programm. Der sicheren Differenzierung der verschiedenen Gefäßsysteme und der Strukturen im Leberhilus wurde hier Priorität gegenüber der Identifikation der Lebersegmente gegeben.

Zum Abschluss wurden 15 Minuten lang pathologische Bilder der besprochenen Organsysteme im Rahmen eines Quiz erarbeitet. Diese Bilder waren aus lizenzrechtlichen Gründen nicht Teil des öffentlich zugänglichen E-Learnings.

### **2.2.2 Zweite Einheit**

Zu Beginn wurden allfällige Fragen oder Unklarheiten der ersten Einheit besprochen und deren Inhalte kurz wiederholt. Dann folgte das praktische Üben der Darstellung der extrahepatischen Gallengänge und der Vena portae im Ligamentum hepatoduodenale via des verlängerten Interkostalschnittes rechts. Sononym für den verlängerten Interkostalschnitt werden häufig auch die Begriffe „CPC-Schnitt“ (Choledochus-Porta-Cava-Schnitt) und „Schulter-Nabel-Schnitt“ verwendet [37]. Anschließend sollten die großen retroperitonealen Gefäße Aorta abdominalis und Vena cava inferior im medianen Längs- und Querschnitt auf verschiedenen Höhen dargestellt werden.

Wie schon zum Abschluss der ersten Einheit wurden erneut Beispielbilder von Pathologien im Rahmen eines Quiz durchbesprochen. Die Proband:innen wurden außerdem daran erinnert, zur letzten Einheit möglichst nüchtern zu erscheinen, um als Schallmodell für die Prüfung geeignet zu sein.

### **2.2.3 Dritte Einheit**

Zu Beginn der dritten Einheit wurde den Proband:innen der konkrete Wortlaut der acht Prüfungsaufgaben sowie die einzelnen Kategorien des Bewertungsschemas offenbart.

Anschließend erfolgte die Darstellung der linken Niere, der Milz sowie des Recessus splenorenalis im Längs- und Querschnitt an der linken Flanke. Den Abschluss bildete die Darstellung der Harnblase und der inneren Geschlechtsorgane des männlichen und weiblichen Beckens im suprasymphysären Längs- und Querschnitt. Auch die peritonealen

Umschlagfalten Recessus rectouterina bei der Frau respektive Recessus rectovesicalis beim Mann wurden identifiziert. Außerdem bestand die Möglichkeit, gezielt die Prüfungsinhalte der beiden vorangegangenen Einheiten zu wiederholen.

In den letzten 30 bis 40 Minuten der dritten Einheit fand die Prüfung statt. Der Ablauf dieser ist in Abschnitt 2.5.3 erläutert.

## **2.3 Kursablauf**

Die Wissensvermittlung fand in drei Einheiten zu je zwei Stunden statt. Zwischen den Einheiten war im Durchschnitt je eine Woche Pause. Die dritte und letzte Einheit, an deren Ende auch die praktische Prüfung stattfand, wurde spätestens 4 Wochen nach dem Ersttermin abgehalten.

Alle Kleingruppen wurden ausschließlich vom Autor der Arbeit persönlich betreut, welcher in der Folge als „Tutor“ bezeichnet wird. Er ist zertifizierter Student:innen-Tutor der ÖGUM (Österreichische Gesellschaft für Ultraschall in der Medizin) im Bereich der Abdomensonographie. Es bestehen zwei Jahre Peer-Teaching-Erfahrung im Rahmen der ehrenamtlichen Tätigkeit bei Sono4you Graz, einer studentischen Initiative, welche in Abschnitt 1.2.4 erläutert wurde. In dieser Studie wurde bewusst nur ein Tutor eingesetzt, um die Vergleichbarkeit zwischen den Gruppen zu wahren und nicht durch unterschiedliche Tutor:innen weitere Heterogenität in das Studiendesign einzubringen.

### **2.3.1 Kontrollgruppe**

In der Kontrollgruppe stellte der Tutor jeweils einen Schnittebene vor, die danach von den Proband:innen praktisch erlernt wurde, bevor der nächste Standardschnitt behandelt wurde. Die theoretische Präsentation der Schnittebenen inklusive Schallkopfposition und anatomischen Gegebenheiten erfolgte anhand des E-Learnings.

Während der Zeiten des praktischen Übens folgte der Tutor einem didaktischen 4-Stufen-Schema bezüglich der Leistung von Hilfestellungen. Zuerst wurde abgewartet, um den Proband:innen genügend Zeit zu geben, selbst eine Frage zu formulieren oder um Hilfe zu bitten. Danach wurden die Proband:innen durch Denkanstöße bei der selbstständigen Problemlösung unterstützt. Dabei wurden sie beispielsweise darauf hingewiesen, sich noch einmal die Anatomie der gesuchten Struktur vor Augen zu führen oder die aktuellen Geräteeinstellungen und die Patient:innenposition zu überdenken. Erst wenn dieses hinweisende Vorgehen nicht zum Erfolg führte, wurden konkrete Handlungsanweisungen

zur Bildoptimierung gegeben. Als ultima ratio wurde physisch in den Prozess der Bildentstehung eingegriffen oder eine spezifische Bewegung vorgeführt. Ein praktisches Vorführen der zu erlernenden Schnitte durch den Tutor war nicht routinemäßig vorgesehen. Abbildung 2 stellt das 4-Stufen-Schema graphisch als Flussdiagramm dar.

Eine weitere Aufgabe des Tutors war es, das Zeitmanagement im Überblick zu behalten und darauf zu achten, dass die Übungszeit annähernd gleich auf die Proband:innen verteilt war. Außerdem animierte er zum Wechsel der Schallmodelle, damit die Proband:innen an möglichst vielen Personen üben. Dies sollte der Gruppenbildung innerhalb der Kleingruppe vorbeugen und die Proband:innen mit verschiedenen anatomischen Situationen konfrontieren.

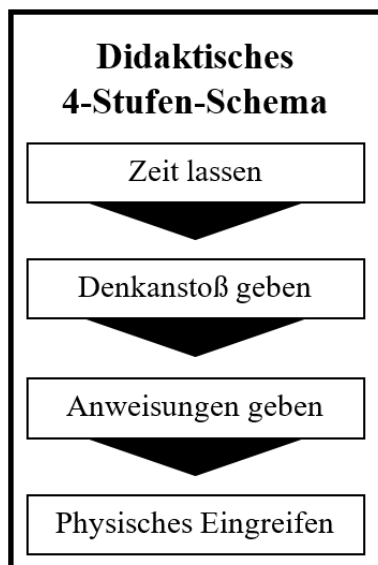


Abbildung 2: Flussdiagramm des didaktischen 4-Stufen-Schemas

### 2.3.2 Interventionsgruppe

Bei der Interventionsgruppe gab es keine theoretische Präsentation der Schnittebenen und kein helfendes Eingreifen des Tutors während der Übungsphasen. Der Tutor war jedoch jederzeit im Raum, um bei technischen Problemen weiterzuhelfen. Das Erlernen der theoretischen und praktischen Fähigkeiten erfolgte rein durch das Selbststudium des bereitgestellten E-Learnings. Bezüglich des Curriculums, der Räumlichkeiten und der Ultraschallgeräte waren die beiden Studiengruppen ident.

Die einführende Phase zu Beginn der ersten Einheit war sehr ähnlich zu der der Kontrollgruppe (siehe Abschnitt 2.3.1). In der Interventionsgruppe wurde sie insbesondere

dazu genutzt, den Proband:innen Möglichkeiten aufzuzeigen, wie ein effektives Selbststudium gelingen könnte. Hier wurde vor allem betont, sich mit der Schallkopfführung, der Gerätekunde und der anatomischen Orientierung am Ultraschallbild vertraut zu machen. Außerdem wurden die häufigsten Problemlösungsstrategien wie das Geben von Atemkommandos sowie die Adaptierung der Position des Schallmodells besprochen. Auch auf die Wichtigkeit des häufigen Wechsels der Schallmodelle innerhalb der Kleingruppe wurde hingewiesen.

Zum Ende jeder Einheit erfolgte, wie auch in der Kontrollgruppe, ein 15-minütiges Pathologie-Quiz, welches vom Tutor moderiert wurde. Da dieses Quiz aus lizenzrechtlichen Gründen nicht im E-Learning veröffentlicht werden konnte, war es nicht im Selbststudium möglich.

## **2.4 Materialien**

### **2.4.1 E-Learning**

Das E-Learning wurde 2021 von vier studentischen Tutor:innen von Sono4you Graz als Vorbereitungsunterlagen für einen Abdomensonographie-Kurs im Flipped-Classroom-Format entwickelt [37]. Bei diesen Kursen bereiten sich die Studierenden routinemäßig mit diesem E-Learning vor, um während der Kurszeit die Nutzung der Übungszeit zu maximieren. Im Sinne des FOAMed-Gedankens (Free Open Access Medical Education) ist das E-Learning öffentlich und kostenlos für alle Personen zugänglich.

Es wurde fachlich von Fachärzt:innen der Fachrichtungen Radiologie sowie Gastroenterologie und Hepatologie validiert. 2021 wurde es mit einem Anerkennungspreis bei der Verleihung des „Ars Docendi Staatspreis für exzellente Lehre an Österreichs öffentlichen Universitäten, Fachhochschulen, Privatuniversitäten und Pädagogischen Hochschulen“ geehrt [38].

Jedes Kapitel erklärt je einen Standardschnitt der Abdomensonographie nach den Richtlinien der DEGUM [2]. Es werden jeweils die Schallkopfposition am Schallmodell, die anatomischen Gegebenheiten sowie die notwendigen Geräteeinstellungen und sonstige spezifische Problemlösestrategien besprochen. An jedes Kapitel schließt ein kurzes Quiz zur Selbstüberprüfung an.

Das E-Learning wurde mit dem Programm „Rise 360“ (Articulate Global LLC, USA) erstellt. Es wurden neben Textelementen auch Grafiken, Fotos, Videos und interaktive Elemente platziert. Die Darstellung des Kurses passt sich der Bildschirmgröße des Endgeräts an. So ist die adäquate Nutzung unterwegs auf einem Tablet oder Smartphone möglich. Die Proband:innen der Interventionsgruppe waren angehalten, ein mobiles Endgerät in Form eines Laptops, Tablets oder Handys mitzubringen, um das E-Learning während der Übungszeiten parallel offen haben zu können.

### **2.4.2 Räumlichkeiten**

Alle Kurse wurden in einem 38,9 m<sup>2</sup> großen Seminarraum der Medizinischen Universität Graz durchgeführt. Es standen ein PC sowie ein Beamer für die Vorträge und das Quiz bereit. Der Raum war zum Schutz der Privatsphäre und für den besseren Kontrast der Graustufen am Display des Ultraschallgeräts mittels Jalousien nahezu vollständig abdunkelbar.

Als Ersatz für die mangelnden Untersuchungsliegen wurden die praktischen Übungseinheiten auf 4 Tischen durchgeführt. Je zwei Tische wurden in einer L-förmigen Konfiguration aufgestellt und bildeten so gemeinsam einen Arbeitsplatz. Darauf kamen jeweils eine Isoliermatte und ein Polster zu liegen.

### **2.4.3 Technische Geräte**

Es wurden 2 Butterfly IQ+ Ultraschallgeräte (Butterfly Network Inc., USA) verwendet. Dabei handelt es sich um mobile Ultraschallsonden, die verschiedene Schallkopfarten imitieren können und somit vielseitig und direkt am Patientenbett einsetzbar sind. Sie können mit diversen Tablets oder Smartphones zur Bildausgabe mittels eines Kabels verbunden werden. In dieser Studie wurden iPads der 8. Generation (Apple Inc., USA) als Ausgabegerät verwendet.

Die Ultraschallgeräte und zugehörigen Tablets wurden von Sono4you Graz 2021 nach einer Förderung durch die „Friedrich Flick Förderungstiftung“ angekauft und können seitdem von allen Studierenden der Medizinischen Universität Graz an der Universitätsbibliothek für je 72 Stunden ausgeliehen werden [39].

## 2.5 Prüfung des Lernerfolges

### 2.5.1 Prüfungsaufgaben

Aus den Lehrinhalten wurden acht Prüfungsaufgaben zusammengestellt. Alle Aufgaben enthielten die klare Anweisung, ein Organ oder eine Struktur darzustellen und das Ultraschallbild entsprechend zu interpretieren. Die Aufgaben wurden so gewählt, dass sie alle einer ähnlichen Schwierigkeitsstufe entsprachen und in einem ähnlichen Zeitrahmen bearbeitbar waren. Der Inhalt aller acht Aufgaben wurde den Proband:innen zu Beginn der dritten Einheit präsentiert, der konkrete Wortlaut ist Tabelle 1 zu entnehmen.

Die Kombination der Aufgaben „Rechte Niere“ und „Linke Niere“ wurde im Vorfeld als nicht zulässig definiert, da sich die Interpretation der beiden Niere zu sehr ähnelt und somit die Trennschärfe in dieser Kategorie zu gering gewesen wäre. Aufgrund der zu großen topographischen Ähnlichkeit wurde auch die Kombination der Aufgaben „Lebervenen“ und „Vena portae“ sowie die Kombination „Linke Niere“ und „Milz“ ausgeschlossen. War die zweite Aufgabe nicht mit der ersten gezogenen Aufgabe kompatibel, so wurde die zweite Aufgabe erneut ausgelost.

Die Darstellung der extrahepatischen Gallengänge sowie der extrahepatischen Vena portae im Ligamentum hepatoduodenale erwies sich im Laufe der Kurse als zu inkonsistent darstellbar bei den nicht-nüchternen Schallmodellen. Aus diesem Grund wurden diese Strukturen bei der Auswahl der Prüfungsaufgaben nicht berücksichtigt. Auch die Durchmusterung der Leber, ein Standardablauf in der Abdomensonographie, wurde nicht als Prüfungsaufgabe eingebaut, da dies viele verschiedene Schnittebenen erfordert hätte und so die Beurteilung im Vergleich zu den übrigen Prüfungsaufgaben verzerrt gewesen wäre.

Auch die Darstellung der Harnblase findet sich nicht unter der finalen Liste an Prüfungsaufgaben. Von der Videodokumentation intimer Bereiche suprasymphysär wurde aus Rücksicht auf die Schallmodelle Abstand genommen. Hinzu kommt, dass das Geschlecht des Schallmodells einen Einfluss auf Bildentstehung und Bildinterpretation hat. Da nicht in jeder Kleingruppe beide Geschlechter vertreten waren, war die nötige Vergleichbarkeit nicht gegeben.

Tabelle 1: Kurzbezeichnung und Wortlaut der Prüfungsaufgaben

<b>Kurzbezeichnung</b>	<b>Wortlaut der Prüfungsaufgabe</b>
<b>Gallenblase</b>	Stellen Sie die Gallenblase zur Gänze in zwei Ebenen (longitudinal und transversal) dar. Beurteilen Sie das Lumen und die Wand. Messen Sie die Wanddicke.
<b>Vena cava inferior</b>	Stellen Sie die Vena cava inferior im Längsschnitt möglichst proximal dar. Beurteilen Sie die Atemvariabilität.
<b>Rechte Niere</b>	Stellen Sie die rechte Niere zur Gänze in zwei Ebenen (longitudinal und transversal) dar. Messen Sie Länge, Breite und Dicke/Tiefe möglichst im Bereich des Nierenhilus. Beurteilen Sie das Nierenparenchym sowie das Nierenbecken-Kelchsystem.
<b>Linke Niere</b>	Stellen Sie die linke Niere zur Gänze in zwei Ebenen (longitudinal und transversal) dar. Messen Sie Länge, Breite und Dicke/Tiefe möglichst im Bereich des Nierenhilus. Beurteilen sie das Nierenparenchym sowie das Nierenbecken-Kelchsystem.
<b>Lebervenen</b>	Stellen Sie die Einmündung der Lebervenen in die Vena cava inferior dar und benennen Sie diese. (Mindestens 2 von 3 Venen)
<b>Vena portae</b>	Stellen Sie die Aufzweigung der Vena portae intrahepatisch dar. Beurteilen Sie die intrahepatischen Gallengänge.
<b>Milz</b>	Stellen Sie die Milz dar. Messen und beurteilen Sie ihre größte Längsausdehnung.
<b>Aorta abdominalis</b>	Stellen Sie die Aorta abdominalis in ihrem Verlauf von subxiphoidal bis zur Bifurcatio aortae im Querschnitt dar. Beurteilen Sie die einsehbaren Areale auf das Vorliegen einer Ektasie oder eines Aneurysmas. Messen Sie an einer Stelle den sagittalen und transversalen Durchmesser.

### 2.5.2 Prüfungszeit

In der Originalarbeit zum OSAUS-Bewertungsschema von Tolsgaard et al. (2013) wurde im Delphi-Konsens beschlossen, dass die benötigte Zeit zwar ein relevantes, aber kein essentielles Kriterium bei der Beurteilung der sonographischen Fähigkeiten darstellt und schnelles, aber unvollständiges Arbeiten nicht mit Punkten belohnt werden soll [23]. Da die Prüfungszeit nicht als Beurteilungselement in das OSAUS-Bewertungsschema Einzug gefunden hat, wurde sie auch in dieser Arbeit nicht gesondert bewertet.

Es besteht jedoch die Hypothese, dass die Prüfungszeit negativ mit den Gesamtpunkten korreliert, da geübtere Proband:innen mit mehr Gesamtpunkten weniger Zeit benötigen als ungeübtere Proband:innen. Insbesondere bei der Kategorie „System“ besteht die Hypothese, dass systematischeres Arbeiten auch rascheres Arbeiten bedeutet und somit die negative Korrelation besonders hoch ausfällt.

Inklusive Vor- und Nachbereitung standen 10 Minuten pro Proband:in zur Verfügung, weswegen das Zeitlimit auf insgesamt 8 Minuten für die Absolvierung von 2 Prüfungsaufgaben festgesetzt wurde. Für die erste Aufgabe standen maximal 4 Minuten zur Verfügung. War nach 4 Minuten die erste Aufgabe nicht vollständig bearbeitet, so wurde diese abgebrochen und zur zweiten Prüfungsaufgabe übergegangen.

### **2.5.3 Prüfungsablauf**

Als Aufsichtsperson und organisatorische Leitung während der Prüfung fungierte der Autor der Studie, welcher auch die Proband:innen tutoriert hatte.

Vor der Prüfung wurde den Proband:innen zwei unterschiedliche Aufgaben zugelost. Sie hatten die freie Wahl, welche Aufgabe sie zuerst bearbeiten wollten. Es wurde aus der Kleingruppe eine Person zufällig als Schallmodell ausgewählt. Vor jeder neuen Prüfung wurden die Bildeinstellungen auf die Voreinstellung „Abdomen“ mit 15 cm Tiefe und 50% Gain zurückgesetzt, um immer gleiche Ausgangsbedingungen zu garantieren.

Die Prüfung begann mit dem Verlesen der ersten Aufgabe, danach wurde die Zeitnehmung gestartet. Die Proband:innen waren angehalten, die Aufgabe vollständig auszuführen und ein oder mehrere Bilder zu dokumentieren. Es wurden keinerlei Hilfestellungen gewährleistet und auch die zu dokumentierenden Bilder mussten die Proband:innen selbstständig wählen. Jede Hilfe von außen, sei es durch Gruppenkolleg:innen oder durch das Nachlesen im E-Learning war untersagt.

Wenn die Proband:innen zu verstehen gaben, dass sie mit der Bearbeitung der Aufgabe fertig waren, wurden sie zur Interpretation des Ultraschallbildes aufgefordert. Produzierten sie auf diese Aufforderung hin keine adäquaten Antworten, wurden konkrete Fragen zur Interpretation gestellt. Nach dem Ende der Prüfung wurden sowohl das Gerät, als auch die Unterlage desinfiziert.

### **2.5.4 Schallmodell**

Als Schallmodell wurde eine Person aus der Kleingruppe zufällig ausgewählt. Im Rahmen des kontrollierten Studiendesigns wurde von der Akquirierung realer Patient:innen mit fassbaren Pathologien Abstand genommen. Da die Fertigkeiten im Hands-on-Training lediglich an jungen, gesunden Personen erlernt wurden, sollte die Prüfung auch nur an eben solchen stattfinden. Diese Studie beschränkte sich auf das Erlernen der abdomensonographischen Basisfähigkeiten. Im Gegensatz dazu untersuchten andere Arbeiten, welche mit realen Patient:innen mit fassbaren Pathologien arbeiteten, die Fähigkeiten von approbierten Ärztinnen und Ärzten [28,29,31].

### **2.5.5 Dokumentation**

Die Handbewegungen der Proband:innen sowie das Abdomen der Schallmodelle wurden mit der Videokamera eines Mobiltelefons (Redmi Note 8T) dokumentiert. Dieses war auf einem Stativ neben dem Tisch montiert und nahm auch den Ton auf.

Allen Proband:innen wurde eine zufällige Identifikationsnummer zugeteilt, welche im Ultraschallbild als Beschriftung eingefügt wurde. Der Output des Ultraschallgerätes wurde am Tablet mittels Bildschirmaufnahme dokumentiert. Die gespeicherten Bilder wurden mit der Identifikationsnummer und der Nummer der Aufgabe beschriftet. Die Zeit pro Prüfungsaufgabe wurde aufgrund der zu geringen Trennschärfe bezüglich Prüfungsbeginn und Prüfungsende auf die nächste halbe Minute gerundet.

Zur Beurteilung wurden das Video der Handykamera und die Bildschirmaufnahme in einem Video-Bearbeitungsprogramm (VideoPad Video-Editor) zusammengefügt, synchronisiert und als Bild-im-Bild-Format neu exportiert. Abbildung 3 zeigt ein Beispielbild aus einer Prüfungssituation.

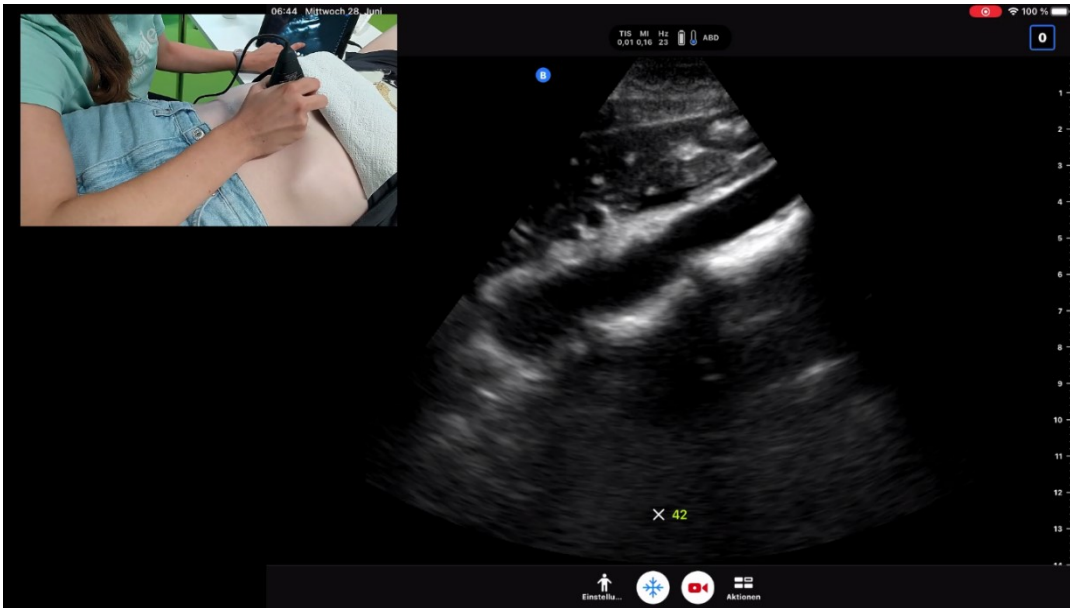


Abbildung 3: Beispielbild aus einer Prüfungssituation (mit freundlicher Genehmigung der Schauspieler:innen)

## 2.5.6 Prüfer:innen

Als Prüfer:innen werden im Folgenden jene vier Personen, zwei Männer und zwei Frauen, bezeichnet, welche die Prüfungsvideos der Proband:innen beurteilten. Alle Prüfer:innen sind ebenso wie der Autor der Studie zertifizierte studentische Tutor:innen der ÖGUM im Bereich der Abdomensonographie und langjährig ehrenamtlich bei der Studierendeninitiative Sono4you Graz tätig. Letztere wird in Abschnitt 1.2.4 erläutert. Alle Prüfer:innen waren nicht anderweitig in die Gestaltung des Studienprotokolls oder die Lehre der Proband:innen involviert.

Die Prüfer:innen waren bei der Bewertung der Videos gegenüber der Studiengruppe der Proband:innen sowie gegenüber den Bewertungen der anderen Prüfer:innen verblindet. Aufgrund der Gegebenheiten der Videoaufnahmen konnten sie hinsichtlich des Geschlechts und der Identität der Proband:innen sowie der Schallmodelle nicht verblindet werden. Um einer Verzerrung vorzubeugen, wurden die Prüfer:innen vorab dazu aufgefordert, persönliche Bekanntschaften bekanntzugeben. 2 Prüfer:innen gaben je eine Person an. In der Folge wurden ihnen jene Videos, in denen diese Personen als Proband:innen oder als Schallmodelle mitwirkten, nicht zur Beurteilung vorgelegt.

Jedes Video wurde von exakt zwei verschiedenen Prüfer:innen beurteilt. Dabei wurde darauf geachtet, die Verteilung der Studiengruppen bei allen vier Prüfer:innen ausgeglichen zu halten.

## **2.6 Bewertungsschema**

In Abschnitt 1.3 wurden bereits die verschiedenen Möglichkeiten der Beurteilung der praktischen Ultraschall-Fähigkeiten vorgestellt. In dieser Arbeit wurde eine modifizierte Version des Bewertungsschema OSAUS („Objective Structured Assessment of Ultrasound Skills“) verwendet, welches von Tosgaard et al. (2013) konzipiert wurde [23]. In einem dreisstufigen modifizierten Delphi-Konsensus einigten sich 60 Expert:innen auf ein Beurteilungsschema mit sieben Kategorien in denen ein bis fünf Punkte auf einer Likert-Skala vergeben werden. Dieses Schema wurde mit der Intention geschaffen, anhand mehrerer klinischer Situationen zu beurteilen, ob Ärzt:innen in der Lage sind, selbstständig Patient:innen zu untersuchen und zu betreuen. Um das Bewertungsschema für die Beurteilung der Fähigkeiten von Studierenden der Humanmedizin anwendbar zu machen, wurden geringe Modifikationen vorgenommen. Allen voran wurde sich auf die Bewertung der Bildaquirierung fokussiert, anstatt die Prüfungsaufgaben in ein klinisches Szenario einzubetten.

Da verschiedene Standardschnitte der Abdomensonographie abgeprüft wurden, wurde sich gegen die Verwendung einer Checkliste als Bewertungsschema entschieden. Hier hätte für jede Prüfungsaufgabe eine eigene Checkliste erarbeitet werden müssen, womit die Vergleichbarkeit der Prüfungsaufgaben nur unzureichend gegeben gewesen wäre.

### **2.6.1 Kategorien**

Die sieben Kategorien nach Tosgaard et al. (2013) lauten:

- (1) Indikation der Untersuchung
- (2) Bedienung des Ultraschallgeräts
- (3) Optimierung des Ultraschallbildes
- (4) Systematik der Untersuchung
- (5) Interpretation der Ultraschallbilder
- (6) Dokumentation der Untersuchung
- (7) Medizinische Entscheidungsfindung anhand der Untersuchungsergebnisse [23]

Die Kategorien eins und sieben sollen nur dann bewertet werden, wenn sie auf das konkrete Fallbeispiel anwendbar sind [23]. Aufgrund des Studienaufbaus mit gesunden Proband:innen und ohne Einbettung der Prüfungsaufgaben in einen medizinischen Kontext sind diese beiden Kategorien nicht anwendbar und wurden dementsprechend nicht beurteilt.

Da die schriftliche Dokumentation der Untersuchung nicht Teil des Curriculums war, wurde auch auf die Kategorie sechs verzichtet.

Die Videos der Teilnehmenden wurden somit von den Prüfer:innen in den vier Kategorien „Bedienung des Ultraschallgeräts“, „Bildoptimierung“, „Systematik der Untersuchung“ und „Interpretation der Bilder“ bewertet. Im Folgenden werden die Kategorien, welche kurz als „Gerät“, „Bild“, „System“ und „Interpretation“ bezeichnet werden, im Detail vorgestellt.

### **2.6.1.1 Bedienung des Ultraschallgeräts**

In der Kategorie „Gerät“ lag das Augenmerk auf der Bedienung des Ultraschallgeräts und dem Umgang mit dem Schallkopf. Versierte Untersucher:innen wurden daran erkannt, dass sie den Schallkopf sicher und mit dem nötigen Feingefühl führten. Insbesondere beim Rotieren des Schallkopfes ließen sich Unterschiede in der Professionalität deutlich machen. In dieser Kategorie sollte außerdem beurteilt werden, ob genug Ultraschallgel verwendet wurde und ob der mit dem Schallkopf ausgeübte Druck adäquat war.

Bei der Gerätebedienung wurde darauf geachtet, ob die Tasten und Einstellungen auf Anhieb gefunden wurden. Außerdem wurde das selbstständige Anhalten des Bildes („freeze“), das Hantieren mit den Messfunktionen sowie das Aussuchen des besten Bildes zur Dokumentation bewertet. Die Verwendung der farbkodierten Doppler-Sonographie war bei keiner der Prüfungsaufgaben zwingend nötig und sollte nur nach gründlicher Überlegung und mit konkreter Fragestellung genutzt werden.

### **2.6.1.2 Bildoptimierung**

In der Kategorie „Bild“ wurde die Qualität der dokumentierten Ultraschallbilder bewertet. Vorrangiges Qualitätskriterium war, ob die in der Prüfungsaufgabe vorgegebenen Strukturen adäquat dokumentiert wurden. Dabei wurde insbesondere auf eine entsprechende Anpassung der Bildeinstellungen Gain und Tiefe geachtet. Außerdem sollte die gesuchte Struktur möglichst zentral am Ultraschallbild abgebildet sein.

Bei den Nieren und der Milz sollte der Hilus am Bild zu sehen sein, um die größtmöglichen Ausdehnungen korrekt zu messen [2]. Die Aorta abdominalis sollte im möglichst kreisrunden Querschnitt dargestellt sein, um den wahren sagittal und transversalen Durchmesser bei normaler Ausrichtung auf den Blutfluss zu messen [2].

Wurde eine der Aufgaben nicht in der vorgegebenen Zeit von 4 Minuten fertiggestellt und kein Ultraschallbild dokumentiert, so war diese Aufgabe in der Kategorie „Bild“ mit 0 Punkten zu bewerten.

### **2.6.1.3 Systematik der Untersuchung**

Als systematisches Vorgehen wurde die Orientierung anhand von Leitstrukturen bei der Suche der geforderten Strukturen verstanden. Im Gegensatz dazu stand das wahllose Suchen ohne erkennbares System. Idealerweise sollte nach dem Aufsetzen des Schallkopfes in dem richtigen Areal zuerst mit groben Bewegungen die Leitstrukturen gesucht werden. Danach sollte die abschließende Feinadjustierung durch kleinere Bewegungen erfolgen. War die zu beurteilende Struktur im Bild zu sehen, sollte diese möglichst im Bild behalten werden und die Konzentration auf die technische Bildoptimierung gelegt werden [2]. Punkteabzug gab es hier beim häufigen Verschwinden der gesuchten Struktur aus dem Ultraschallbild und zahlreichen Neuorientierungen.

Es wurden außerdem die Problemlösestrategien der Proband:innen, beispielsweise anhand der adäquaten Gabe von Atem- und Lagerungskommandos, beurteilt. Bei der Abdomensonographie ist es häufig sinnvoll, das Schallmodell je nach Fragestellung in Links- oder Rechtsseitenlage zu bringen. Zusätzlich kann das Elevieren des Arm über den Kopf die interkostalen Räume aufspreizen und so Rippenschatten umgehen [2]. Atemkommandos sind insbesondere zur vollständigen Beurteilung der Oberbauchorgane meist unerlässlich. Gegebenenfalls ist das Ausweichen auf einen intercostalen Schnitt nötig [2].

Auch die Vollständigkeit der Durchmusterung der gefragten Struktur wurde bewertet. Das vollständige Durchfächern ist notwendig, um etwaige Pathologien an den Organrändern nicht zu übersehen.

### **2.6.1.4 Interpretation der Bilder**

In der Kategorie „Interpretation“ wurden Punkte für die Beschreibung des vorliegenden Ultraschallbefundes unter Verwendung einer angemessenen Fachsprache verwendet. Wenn anwendbar, floss auch die korrekte Einordnung der vorgenommenen Messungen sowie die korrekte Benennung der anatomischen Strukturen in die Beurteilung ein.

Da bei allen Schallmodellen physiologische Verhältnisse vorlagen, wurden die Proband:innen dazu aufgefordert zu erläutern, auf welche Weise sie Pathologien ausgeschlossen haben und wie sich diese präsentiert hätten.

### **2.6.2 Punktevergabe**

Im von Tolsgaard et al. (2013) publizierten Bewertungsschema OSAUS war die Vergabe von ein bis fünf Punkten pro Kategorie auf einer Likert-Skala vorgesehen [23]. Dies wurde auch in zahlreichen Folgestudien so gehandhabt [17,28,31,40].

In der vorliegenden Studie wurde die Skala auf null bis zehn Punkte pro Kategorie ausgedehnt, um eine feinere Differenzierung zu ermöglichen. Null Punkte in einer Kategorie bedeutete, dass keinerlei Kompetenz ersichtlich war. Sechs Punkte sollten vergeben werden, wenn die Leistung der Proband:innen als gerade noch zufriedenstellend beurteilt wurde. Die Maximalanzahl von zehn Punkten sollte jenen Proband:innen vorbehalten bleiben, deren Niveau über dem lag, was durchschnittlich nach Absolvierung eines Abdomensonographie-Kurses zu erwarten wäre. Die Vergabe von Teilpunkten war ebenso wie die zusätzliche Beurteilung durch Freitextkommentare nicht vorgesehen. Die Punkte wurden nicht gesondert pro Aufgabe erfasst, sondern in Zusammenschau beider Aufgaben mit gleicher Gewichtung vergeben. Dabei waren die Kategorien möglichst unabhängig von einander zu bewerten. Alle Punkte der vier Kategorien wurden gleich gewichtet addiert, sodass die Proband:innen 0 bis 40 Gesamtpunkte erreichen konnten.

Die Bestehensgrenze wurde auf 60% der maximal erreichbaren Punkte festgesetzt, entsprechend einer Mindestanzahl von 24 Gesamtpunkten. Dies wurde so gewählt, da 60% dem üblichen Bestehensniveau bei Fachprüfungen an der Medizinischen Universität Graz entsprach.

### **2.6.3 Kalibrierungsvideos**

Um das Bewertungsschema zu testen und die Prüfer:innen darauf einzuschulen wurden vier Testvideos zur Kalibrierung aufgenommen, welche von den Prüfer:innen unabhängig von einander bewertet wurden. Davor fand eine Schulung der Prüfer:innen statt, in der das Bewertungsschema vorgestellt wurde. Nach der Bewertung der Kalibrierungsvideos erfolgte eine Diskussion der Ergebnisse mit den Prüfer:innen. Als Konsequenz dieser Diskussion wurde das Bewertungsschema zur oben vorgestellten finalen Version geringfügig überarbeitet.

Für die Produktion der Kalibrierungsvideos wurden Proband:innen auf vier verschiedenen Fähigkeitsstufen nachgeahmt. Sie werden in der Folge in absteigender Reihenfolge als „Sehr gut“, „Gut“, „Mittel“ und „Schlecht“ bezeichnet. Die Situation war bezüglich Prüfungsaufgaben, Zeitlimit und allen anderen Faktoren ident zu der Prüfungssituation der Proband:innen. Als Schauspieler:innen fungierten zwei erfahrene Tutor:innen von Sono4you Graz, die dieselben Voraussetzungen wie die Prüfer:innen erfüllten, welche in Abschnitt 2.5.6 erläutert wurden. Außerhalb der Produktion dieser Kalibrierungsvideos waren die Schauspieler:innen nicht anderweitig in die Verwirklichung der Studie involviert.

### **2.6.3.1 Niveau der Videos**

Im ersten Kalibrierungsvideo „Sehr gut“ wurde ein:e Proband:in auf sehr gutem Niveau, ähnlich dem einer Tutorin oder eines Tutors gemietet. Es wurden die Prüfungsaufgaben „Vena portae“ und „Aorta abdominalis“ behandelt. Als Zielwert wurden 85 bis 95% der maximal erreichbaren Punkte definiert.

Im zweiten Kalibrierungsvideo „Gut“ sollte ein fortgeschrittenes Niveau erkennbar sein. Es grenzte sich vom sehr guten Video durch Unsicherheit, mangelndes vollständiges Durchfächern der Strukturen sowie fehlende Problemlösestrategien ab. Die Schallkopfführung und die Bildeinstellungen imponierten jedoch merklich geübter als in den „mittleren“ und „schlechten“ Videos. Es wurden die Prüfungsaufgaben „Lebervenen“ und „Vena cava inferior“ behandelt. Bei den Lebervenen konnten nur zwei von drei Gefäßen dargestellt werden, was die fälschliche Benennung nach sich zog. Bei der Darstellung der Vena cava inferior bereitete anfangs die Abgrenzung des Gefäßes von der Aorta abdominalis Schwierigkeiten. Als Zielwert wurden 70 bis 80% der maximal erreichbaren Punkte definiert.

Im dritten Kalibrierungsvideo „Mittel“ wurde ein:e Proband:in durchschnittlichen Niveaus verkörpert. Anweisung an die Schauspieler:innen waren fehlende Atemkommandos und mangelnde anatomische Intuition der Schallkopfbewegungen sowie keine ausreichende Orientierung an Leitstrukturen. Es wurden die Prüfungsaufgaben „Gallenblase“ und „Linke Niere“ behandelt. Bei der Darstellung der Gallenblase erfolgte der unreflektierte Einsatz des farbkodierten Duplexsonographie mit nachfolgender Fehlinterpretation. Die Gallenblasewand wurde leberfern anstatt korrekterweise lebernah gemessen [2]. Bei der linken Niere gelang keine eindeutige Abgrenzung eines Längs- und Querschnittes, sondern lediglich eine querovale Darstellung. Als Zielwert wurden 55 bis 65% der maximal

erreichbaren Punkte definiert. Es sollte somit ein Niveau nahe an der Bestehensgrenze dargestellt werden.

Im abschließenden vierten Kalibrierungsvideo „Schelcht“ wurde ein:e Proband:in gemietet, der:die definitiv nicht bestehen sollte. Es waren kaum Fähigkeiten in allen Kategorien erkennbar. Sowohl das Hantieren mit dem Schallkopf und dem Ultraschallgerät als auch die Bildeinstellungen und die Kommunikation mit dem Schallmodell wirkten laienhaft. Als Zielwert wurden 10 bis 20% der maximal erreichbaren Punkte definiert.

### **2.6.3.2 Ergebnisse**

Analog zu Canty et al. (2019) wurde eine Abweichung von unter 30% des Mittelwerts der Gesamtpunkte eines Videos als eine akzeptable Übereinstimmung zwischen den Prüfer:innen definiert [16]. Es zeigte sich, dass Prüfer:in A bei drei von vier Kalibrierungsvideos weniger Punkte als diese Untergrenze vergab, wohingegen Prüfer:in C bei zwei Videos über der Obergrenze lag. Mit diesen beiden Prüfer:innen wurden die Videos gesondert durchgearbeitet und deren Argumentationen angehört, bis ein Konsens geschlossen werden konnte und das Bewertungsschema entsprechend angepasst war.

Beim Vergleich der Mittelwerte der Gesamtpunkte mit den vorab definierten Zielwerten zeigte sich, dass zwei von vier Videos niedriger als der Zielbereich bewertet wurden. In Zusammenschau mit den Prüfer:innen bestand der Konsens, dass die Imitationen der Niveaus bei beiden Videos Limitationen hatten und keine systematischen Fehler der Prüfer:innen vorlagen. Tabelle 2 zeigt eine Übersicht über die von den Prüfer:innen vergebenen Punkte.

Tabelle 2: Bewertung der Kalibrierungsvideos

Niveau		A	B	C	D	Gesamtpunkte (M ± SD)	± 30% (von M)	Zielwert
<b>Sehr gut</b>	Punkte	30	27	33	37	31,8 (± 3,7)	22,3 – 41,3	34 – 38
	Prozent					79,5%		85 – 95%
<b>Gut</b>	Punkte	19	30	36	35	30,0 (± 6,7)	21,0 – 39,0	28 – 32
	Prozent					75,0%		70 – 80%
<b>Mittel</b>	Punkte	6	20	29	18	18,3 (± 8,2)	12,8 – 23,8	22 – 26
	Prozent					45,8%		55 – 65%
<b>Schlecht</b>	Punkte	2	7	17	5	7,8 (± 5,6)	5,5 – 10,1	4 – 8
	Prozent					19,5%		10 – 20%

(A-D = Prüfer:innen, M = Mittelwert, SD = Standardabweichung)

## 2.7 Statistik

Es wurde eine a priori Poweranalyse durchgeführt. Bei einer erwarteten Überlegenheit der Kontrollgruppe mit einer Effektgröße von  $d = 1,0$ , einem Fehler 1. Art von  $\alpha = 0,05$  und einer Power von  $\beta = 0,90$  waren 23 Proband:innen pro Gruppe erforderlich. Die Drop-out-Rate wurde nach Erfahrungswerten bei der Durchführung ehrenamtlicher, freiwilliger extracurriculärer Kurse bei Sono4you Graz auf 5,0% geschätzt. Dementsprechend wurde eine Proband:innenanzahl von 50 Personen angepeilt. Als Obergrenze wurden aus räumlichen und zeitlichen Kapazitätsgründen 60 Personen definiert, entsprechend 10 Kleingruppen zu je sechs Proband:innen. Der didaktisch minimal relevante Unterschied wurde analog zu einer Vorarbeit auf vier Gesamtpunkte, entsprechend 10 Prozentpunkten, geschätzt [29].

Die statistische Analysen wurden mit SPSS Statistics 28 (IBM Corporation, USA) durchgeführt. Sämtliche statistischen Analysen wurden auf dem zweiseitigen Signifikanzniveau  $\alpha = 0,05$  mit ungerichteten Hypothesen durchgeführt, um eine mögliche Überlegenheit der Interventionsgruppe zu berücksichtigen. Für den Hauptzielparameter der Gesamtpunkte in Abhängigkeit der Studiengruppe wurde zusätzlich die einseitige Signifikanz berechnet, da die gerichtete Hypothese bestand, dass die Kontrollgruppe mit einer Effektstärke von  $d = 1,0$  besser abschneidet als die Interventionsgruppe. Auf eine Korrektur für multiple Vergleiche wurde bei dieser explorativen Datenanalyse verzichtet.

## **2.8 Ethik**

Laut der Stellungnahme der Ethikkommission der Medizinischen Universität Graz vom 21.03.2023 war keine formelle Begutachtung durch eine Ethikkommission notwendig, da es sich um eine Lehr-Evaluation ohne Beteiligung von Patient:innen handle. Es liege keine medizinische Forschung am Menschen im Sinne der Deklaration von Helsinki vor. Bezüglich dem Vorgehen beim Auffinden von Zufallsbefunden bei den Proband:innen wurde nach den Handlungsempfehlungen von Dietrich et al. (2022) vorgegangen [41].

## 3 Ergebnisse

### 3.1 Fallzahl und Drop-out-Rate

Von den 48 randomisierten Proband:innen nahmen 43 an der ersten Einheit Teil (89,6%). 41 Personen davon besuchten die zweite Einheit (95,3%). 39 Personen davon schlossen die Studie zur Gänze ab (95,1%). Nach der dritten Einheit gab es kein weiteres Drop-out, es waren alle Videos technisch einwandfrei und konnten uneingeschränkt beurteilt werden.

Es beendeten somit 9 von 48 Proband:innen die Studie frühzeitig, entsprechend einer Drop-out-Rate von 18,8% in der gesamten Studienpopulation. In der Interventionsgruppe stiegen 3 von 23 Proband:innen (13,0%) aus, in der Kontrollgruppe brachten 6 von 25 Proband:innen (24,0%) die Studie nicht zu Ende. Tabelle 4 zeigt die entsprechende Kreuztabelle. Die exakten Zeitpunkte des Drop-outs der Proband:innen sind im Flussdiagramm in Abbildung 4 nachzulesen.

Aufgrund der Stichprobengröße ( $n = 48$ ), dem einen Freiheitsgrad ( $df = 1$ ) und den Erwartungswerten (2 Zellen mit einem Erwartungswert  $< 5$ ) wurde der exakte Test nach Fisher zur Kontingenzanalyse gewählt [42]. Hier zeigte sich, dass kein signifikanter Unterschied in der Drop-out-Rate zwischen den beiden Studiengruppen bestand ( $p = 0,466$ , zweiseitig).

Tabelle 3: Drop-out nach der Randomisierung in Abhängigkeit der Studiengruppe

		Intervention	Kontrolle	Studienpopulation
<b>Drop-out</b>	Anzahl	3	6	9
	Prozent	13,0%	24,0%	18,8%
<b>Abgeschlossen</b>	Anzahl	20	19	39
	Prozent	87,0%	76,0%	81,2%
<b>Summe</b>	Anzahl	23	25	48
	Prozent	100%	100%	100%

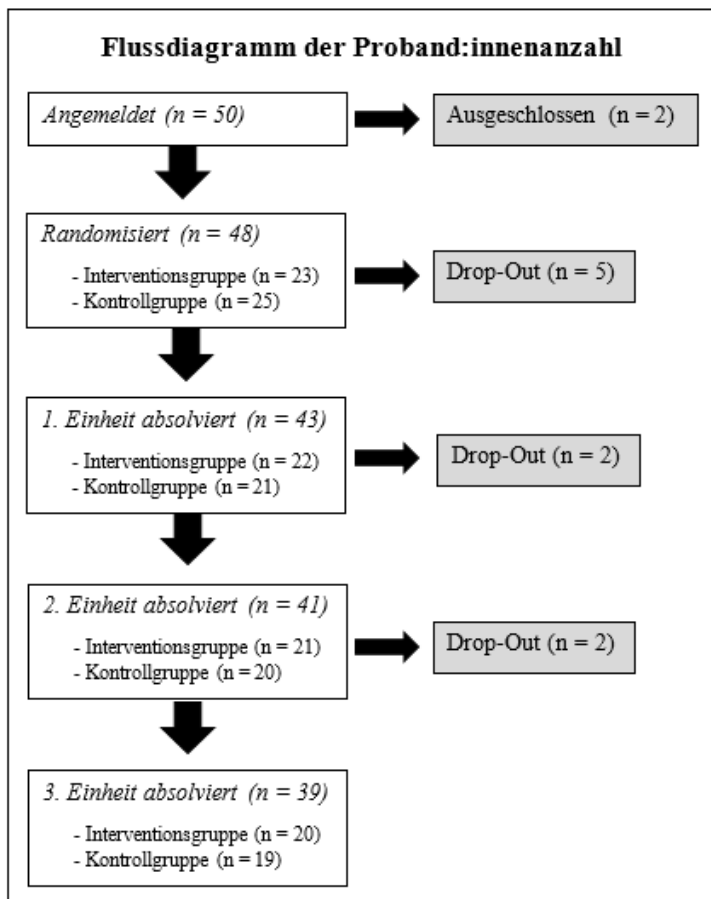


Abbildung 4: Flussdiagramm der Proband:innenanzahl

## 3.2 Demographische Daten

### 3.2.1 Geschlecht

Insgesamt schlossen 14 Männer (35,9%) und 25 Frauen (64,1%) die Studie ab. Keine der Personen hatte sich als divers bezeichnet. Innerhalb der Interventionsgruppe war die Verteilung mit 11 Männern (55,0%) und 9 Frauen (45,0%) annähernd ausgeglichen, wogegen in der Kontrollgruppe die 16 Frauen (84,2%) gegenüber den drei Männern (15,8%) in der Mehrheit waren. Tabelle 5 zeigt die entsprechende Kreuztabelle. In Abbildung 5 sind die Häufigkeiten in einem Balkendiagramm graphisch dargestellt.

Zur Analyse der Unterschiede der Häufigkeiten der Geschlechter in den beiden Studiengruppen wurde die Kontinuitätskorrektur nach Yates angewandt. Diese wurde aufgrund der Stichprobengröße ( $n = 39$ ), dem einen Freiheitsgrad ( $df = 1$ ) und den Erwartungswerten (Keine Zelle mit einem Erwartungswert  $< 5$ ) gewählt [42]. Hier ergab

sich eine statistisch signifikante Ungleichverteilung der Geschlechter auf die beiden Studiengruppen (Kontinuitätskorrektur = 4,918;  $p = 0,027$ ; asymptotisch zweiseitig).

Tabelle 4: Häufigkeiten der Geschlechter in Abhängigkeit der Studiengruppe

		Intervention	Kontrolle	Studienpopulation
<b>Männlich</b>	Anzahl	11	3	14
	Prozent	55,0%	15,8%	35,9%
<b>Weiblich</b>	Anzahl	9	16	25
	Prozent	45,0%	84,2%	64,1%
<b>Summe</b>	Anzahl	20	19	39
	Prozent	100%	100%	100%



Abbildung 5: Häufigkeit der Geschlechter in den Studiengruppen

### 3.2.2 Studienfortschritt

Insgesamt nahmen 22 Personen (56,4%) aus dem 3. Studienjahr, 12 Personen (30,8%) aus dem 4. Studienjahr, vier Personen (10,3%) aus dem 5. Studienjahr und eine Person aus dem 6. Studienjahr (2,6%) an der Studie teil. In Tabelle 6 sind die Häufigkeiten nach Studiengruppen aufgeschlüsselt und in Abbildung 6 in einem Balkendiagramm graphisch dargestellt.

Zum Ausschluss, dass der Studienfortschritt der Proband:innen ungleichmäßig auf die Studiengruppen verteilt war, wurde eine Kontingenzanalyse durchgeführt. Aufgrund der

Stichprobengröße ( $n = 39$ ), dem einen Freiheitsgrad ( $df = 1$ ) und der Erwartungswerte (4 Zellen mit einem Erwartungswert  $< 5$ ) wurde der exakte Test nach Fisher gewählt [42]. Dieser zeigte keinen signifikanten Unterschied in der Verteilung des Studienfortschritts zwischen den beiden Studiengruppen (Teststatistik = 1,395;  $p = 0,891$ ; zweiseitig).

Tabelle 5: Studienfortschritt (in Jahren) der Proband:innen in Abhängigkeit der Studiengruppe

Studienfortschritt		Intervention	Kontrolle	Studienpopulation
<b>3. Studienjahr</b>	Anzahl	11	11	22
	Prozent	55,0%	55,0%	56,4%
<b>4. Studienjahr</b>	Anzahl	7	5	12
	Prozent	35,0%	26,3%	30,8%
<b>5. Studienjahr</b>	Anzahl	2	2	4
	Prozent	10,0%	10,5%	10,3%
<b>6. Studienjahr</b>	Anzahl	0	1	1
	Prozent	0%	5,3%	2,6%
<b>Summe</b>	Anzahl	20	19	39
	Prozent	100%	100%	100%

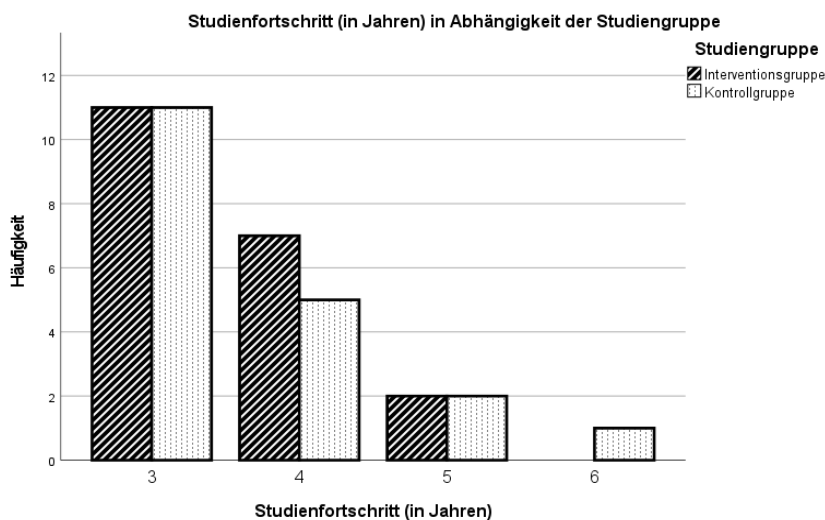


Abbildung 6: Studienfortschritt in den Studiengruppen

### 3.2.3 Gruppengröße

In der gesamten Studienpopulation absolvierten 9 Proband:innen (23,1%) die Studie in Dreiergruppen, 20 Proband:innen (51,3%) in Vierergruppen und 10 Proband:innen (25,6%) in Fünfergruppen. Die Verteilung der Proband:innen in Bezug auf die Gruppengröße und in

Abhängigkeit der Studiengruppe können Tabelle 7 entnommen werden. In Abbildung 7 sind die Häufigkeiten in einem Balkendiagramm graphisch dargestellt.

Zum Ausschluss, dass ein signifikanter Unterschied in den Gruppengrößen zwischen den beiden Studiengruppen bestand, wurde eine Kontingenzanalyse durchgeführt. Aufgrund der Stichprobengröße ( $n = 39$ ), dem einen Freiheitsgrad ( $df = 1$ ) und der Erwartungswerte (3 Zellen mit einem Erwartungswert  $< 5$ ) wurde der exakte Test nach Fisher gewählt [42]. Dieser zeigte keinen signifikanten Unterschied in der Verteilung der Gruppengrößen zwischen den beiden Studiengruppen (Teststatistik = 1,771;  $p = 0,447$ ; zweiseitig).

Tabelle 6: Anzahl der Proband:innen in Kleingruppen mit verschiedener Gruppengröße

Gruppengröße		Intervention	Kontrolle	Studienpopulation
<b>3 Personen</b>	Anzahl	3	6	9
	Prozent	15,0%	31,6%	23,1%
<b>4 Personen</b>	Anzahl	12	8	20
	Prozent	60,0%	42,1%	51,3%
<b>5 Personen</b>	Anzahl	5	5	10
	Prozent	25,0%	26,3%	25,6%
<b>Summe</b>	Anzahl	20	19	39
	Prozent	100%	100%	100%

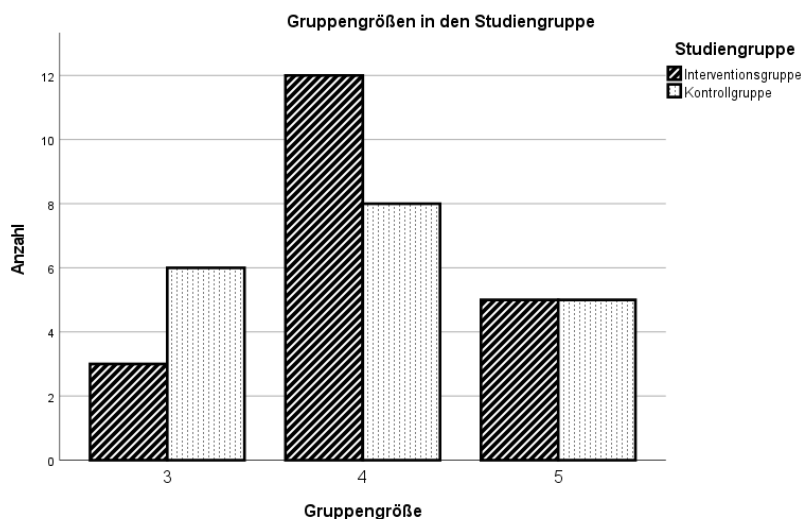


Abbildung 7: Gruppengrößen in den Studiengruppen

### 3.3 Hauptzielparameter

#### 3.3.1 Deskriptive Statistik

In der gesamten Studienpopulation ( $n = 39$ ) wurden im Durchschnitt  $28,9 (\pm 6,3)$  Gesamtpunkte erreicht, entsprechend  $72,2 \%$  der maximal erreichbaren Punkte. Die höchste Zahl von  $37,5$  Gesamtpunkten ( $93,8\%$  der erreichbaren Punkte) erreichte ein:e Proband:in aus der Kontrollgruppe. Am schlechtesten schnitt ein:e Proband:in aus der Interventionsgruppe mit  $11,0$  Punkten ( $27,8\%$  der erreichbaren Punkte) ab.

Die Proband:innen der Interventionsgruppe ( $n = 20$ ) erreichten im arithmetischen Mittel  $27,6 (\pm 6,3)$  Gesamtpunkte, in der Kontrollgruppe ( $n = 19$ ) waren es durchschnittlich  $30,3 (\pm 6,2)$  Gesamtpunkte. Dies entspricht  $69,0\%$  respektive  $75,8\%$  der maximal erreichbaren Punkte. In Tabelle 8 sind die Punkte nach Kategorien und in Abhängigkeit der Studiengruppe aufgeschlüsselt, Abbildung 8 stellt diese graphisch in Boxplots dar.

Tabelle 7: Punkte in Abhängigkeit der Studiengruppe

	<b>Intervention</b> ( $n = 20$ , $M \pm SD$ )	<b>Kontrolle</b> ( $n = 19$ , $M \pm SD$ )	<b>Studienpopulation</b> ( $n = 39$ , $M \pm SD$ )
<b>Gerät</b>	$7,2 (\pm 1,2)$	$7,7 (\pm 1,1)$	$7,5 (\pm 1,2)$
<b>Bild</b>	$6,3 (\pm 2,0)$	$7,3 (\pm 1,7)$	$6,8 (\pm 1,9)$
<b>System</b>	$6,9 (\pm 1,7)$	$7,4 (\pm 1,8)$	$7,1 (\pm 1,7)$
<b>Interpretation</b>	$7,2 (\pm 1,8)$	$7,8 (\pm 1,9)$	$7,5 (\pm 1,8)$
<b>Gesamtpunkte</b>	$27,6 (\pm 6,3)$	$30,3 (\pm 6,2)$	$28,9 (\pm 6,3)$

*n = Proband:innenanzahl, M = Mittelwert, SD = Standardabweichung*

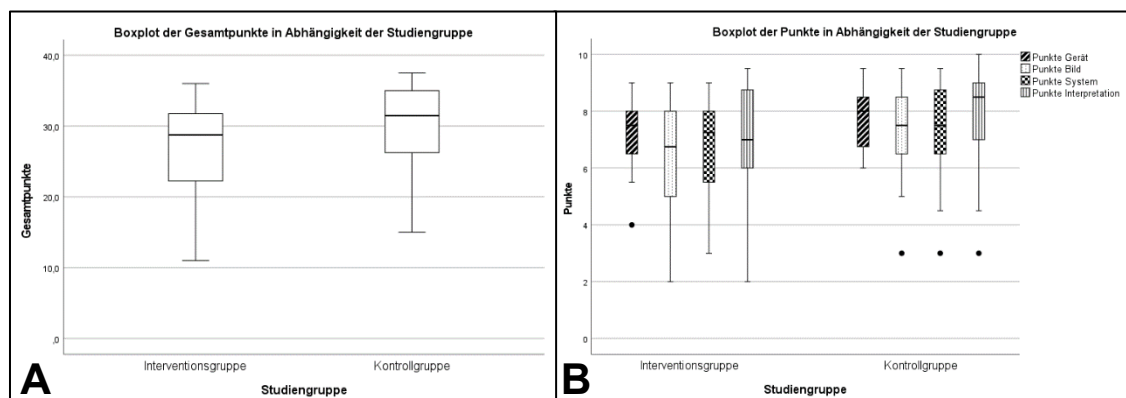


Abbildung 8: Boxplot der Gesamtpunkte (A) bzw. der Punkte in den Kategorien (B) in Abhängigkeit der Studiengruppe

### 3.3.2 Unterschiede zwischen den Studiengruppen

Da die Punkte im Shapiro-Wilk-Test nicht in allen Kategorien als normalverteilt angenommen werden konnten, wurde zur Dependenzanalyse der Mann-Whitney-U-Test gewählt [42]. Aufgrund der a priori Hypothese, dass die Proband:innen in der Kontrollgruppe mehr Punkte erreichen würden, wurde zunächst einseitig getestet. Hier zeigte sich beim Vergleich der Mittelwerte der Gesamtpunkte keine signifikante Überlegenheit der Kontrollgruppe ( $U = 134,500$ ;  $z = -1,561$ ;  $p = 0,061$ ; einseitig) und die Hypothese musste verworfen werden. Auch bei der zweiseitigen Testung, um eine mögliche Überlegenheit der Interventionsgruppe zu berücksichtigen, zeigten sich keine signifikanten Unterschiede zwischen den Studiengruppen ( $p = 0,120$ ; zweiseitig).

Bei den Mittelwerten der Punkte in den Kategorien gab es ebenfalls keinen signifikanten Unterschied zwischen den beiden Studiengruppen (jeweils  $p > 0,05$ ). Teststatistiken und Signifikanzen dazu sind Tabelle 9 zu entnehmen.

Tabelle 8: Ergebnisse des Mann-Whitney-U-Tests der Unterschiede der Punkte zwischen den Studiengruppen

	<b>Teststatistik</b>	<b>Standardisierte Teststatistik</b>	<b>Signifikanz (einseitig)</b>	<b>Signifikanz (zweiseitig)</b>
<b>Gerät</b>	$U = 142,000$	$z = -1,361$	$p = 0,089$	$p = 0,184$
<b>Bild</b>	$U = 133,000$	$z = -1,613$	$p = 0,054$	$p = 0,113$
<b>System</b>	$U = 152,500$	$z = -1,059$	$p = 0,148$	$p = 0,296$
<b>Interpretation</b>	$U = 141,000$	$z = -1,386$	$p = 0,085$	$p = 0,175$
<b>Gesamtpunkte</b>	$U = 134,500$	$z = -1,561$	$p = 0,061$	$p = 0,120$

### 3.3.3 Bestehensquote

Zum Bestehen der Prüfung wurden mindestens 24 Punkte benötigt, entsprechend 60% der Maximalpunktezahl. Dieser Grenzwert wurde in Abschnitt 2.6.2 festgelegt. Insgesamt erreichten 29 der 39 Proband:innen (74,4%) mindestens 24 Punkte. In der Interventionsgruppe bestanden 14 von 20 (70,0%) Proband:innen, in der Kontrollgruppe waren es 15 von 19 Proband:innen (78,9%). Tabelle 10 zeigt die entsprechende Kreuztabelle und Abbildung 9 stellt die Bestehensquoten in den Studiengruppen in einem Balkendiagramm graphisch dar.

Zur Analyse des Zusammenhangs zwischen dem Prozentsatz der bestandenen Proband:innen und der Studiengruppe wurde aufgrund der Stichprobengröße ( $n = 39$ ), der Freiheitsgrade ( $df = 1$ ) und der Erwartungswerte (1 Zelle mit einem Erwartungswert  $< 5$ ) der exakte Test nach Fisher gewählt [42]. Hier zeigte sich kein signifikanter Unterschied in dem Prozentsatz der bestandenen Proband:innen zwischen den Studiengruppen ( $p = 0,716$ ; zweiseitig).

Tabelle 9: : Bestehensquote in Abhängigkeit der Studiengruppe

		Intervention	Kontrolle	Studienpopulation
<b>Bestanden</b>	Anzahl	14	15	29
	Prozent	70,0%	78,9%	74,4%
<b>Nicht Bestanden</b>	Anzahl	6	4	10
	Prozent	30,0%	21,1%	25,6%
<b>Summe</b>	Anzahl	20	19	39
	Prozent	100%	100%	100%

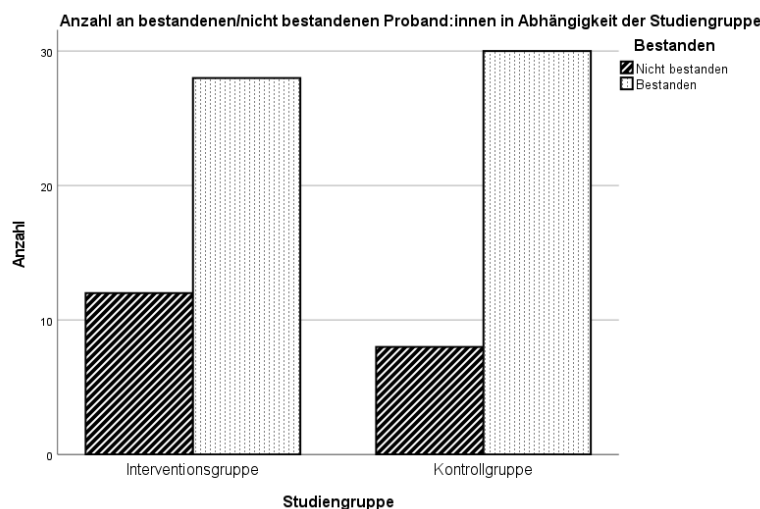


Abbildung 9: Bestanden und nicht bestanden Proband:innen in Abhängigkeit der Studiengruppe

### 3.4 Geschlecht

Die Männer erreichten im Mittel  $27,4 (\pm 7,1)$  Punkte, während die Frauen im Mittel  $29,7 (\pm 5,8)$  Punkte sammeln konnten. Dies entspricht  $68,5\%$  respektive  $74,3\%$  der maximal erreichbaren Punkte. Abbildung 10 stellt die Gesamtpunkte in Abhängigkeit des Geschlechts graphisch mittels eines Boxplots dar.

Ein Shapiro-Wilk-Test zeigte eine Normalverteilung der Gesamtpunkte sowohl bei den Männern ( $p = 0,275$ ), als auch bei den Frauen ( $p = 0,106$ ). Da eine Normalverteilung der Daten angenommen werden konnte und der Levene-Test die Varianzgleichheit ( $p = 0,341$ ) bestätigte, wurde zur Dependenzanalyse der T-Test für unabhängige Stichproben gewählt [42]. Hier unterschieden sich die Mittelwerte der Gesamtpunkte der beiden Geschlechter nicht signifikant voneinander ( $t = -1,137$ ;  $p = 0,263$ ; zweiseitig). Ebenso unterschieden sich die Mittelwerte der Punkte in den einzelnen Kategorien zwischen den Geschlechtern nicht signifikant voneinander (jeweils  $p > 0,05$ ). Die Punkte in Abhängigkeit des Geschlechts und die Ergebnisse der T-Tests sind in Tabelle 11 aufgeschlüsselt.

Tabelle 10: Punkte in Abhängigkeit des Geschlechts sowie Ergebnisse des T-Tests für unabhängige Stichproben

	<b>Männlich</b> (M ± SD)	<b>Weiblich</b> (M ±SD)	<b>Teststatistik</b>	<b>Signifikanz</b> (zweiseitig)
<b>n</b>	14	25		
<b>Gerät</b>	7,2 (± 1, 4)	7,6 (± 1,1)	t = -0,967	p = 0,377
<b>Bild</b>	6.3 (± 2,1)	7,1 (±1,8)	t = -1,202	p = 0,262
<b>System</b>	6,7 (± 1,8)	7,4 (± 1,6)	t = -1,426	p = 0,181
<b>Interpretation</b>	7,2 (± 2,1)	7,6 (± 1,7)	t = -0,685	p = 0,529
<b>Gesamtpunkte</b>	27,4 (± 7.1)	29,7 (± 5,8)	t = -1,137	p = 0,296

*n = Proband:innenanzahl, M =Mittelwert, SD = Standardabweichung*

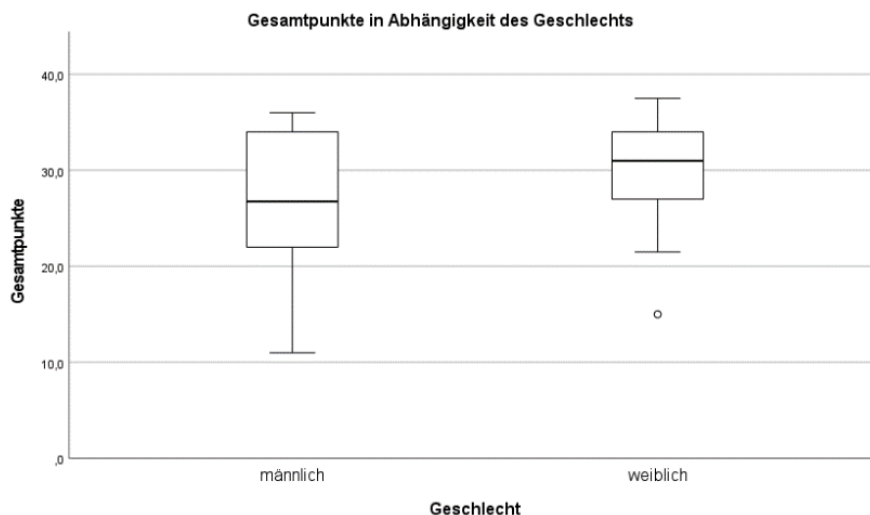


Abbildung 10: Boxplot der Gesamtpunkte in Abhängigkeit des Geschlechts

### 3.5 Studienfortschritt

Die Proband:innen im 3. Studienjahr ( $n = 22$ ) erreichten im Mittel 26,9 ( $\pm 6,5$ ) Punkte, jene im 4. Studienjahr ( $n = 12$ ) erreichten 31,0 ( $\pm 5,3$ ) Punkte und jene im 5. Studienjahr ( $n = 4$ ) erreichten 31,5 ( $\pm 5,3$ ) Punkte. Die Person im 6. Studienjahr ( $n = 1$ ) erreichte 36,0 Punkte. Dies entspricht 67,3%, 77,5%, 78,8% und 90,0% der maximal erreichbaren Punkte.

Insbesondere die minimal erreichten Gesamtpunkte differierten bei den Proband:innen mit unterschiedlichen Studienfortschritt deutlich. Sie erreichten im 3. Studienjahr mindestens 11,0 Punkte, im 4. Studienjahr mindestens 21,0 Punkte und im 5. Studienjahr mindestens 25,5 Punkten. Aus dem 6. Studienjahr nahm nur eine Probandin teil, welche 36,0 Punkte erreichte. Die Daten sind in Tabelle 12 aufgeschlüsselt.

Tabelle 11: Gesamtpunkte in Abhängigkeit des Studienfortschritts

Studienfortschritt	n	Gesamtpunkte (M $\pm$ SD)	Gesamtpunkte (Minimum)	Gesamtpunkte (Maximum)
3. Jahr	22	26,9 ( $\pm 6,5$ ) 67,3%	11,0	37,0
4. Jahr	12	31,0 ( $\pm 5,3$ ) 77,5%	21,0	37,0
5. Jahr	4	31,5 ( $\pm 5,3$ ) 78,8%	25,5	37,5
6. Jahr	1	36,0 ( $\pm 0,0$ ) 90,0%	36,0	36,0

*n = Proband:innenanzahl, M = Mittelwert, SD = Standardabweichung*

Im Shapiro-Wilk-Test zeigte sich, dass die Gesamtpunkte nicht in allen Gruppen als normalverteilt angenommen werden konnten. Bei Annahme einer linearen Korrelation auf dem Streudiagramm (siehe Abbildung 11) und fehlender Normalverteilung der Daten wurde zur Dependenzanalyse die Rangkorrelationsanalyse nach Spearman gewählt [42]. Es zeigte sich eine signifikante positive Korrelation zwischen dem Studienfortschritt und den Gesamtpunkten mit einem Korrelationskoeffizient Rho von  $\rho = 0,39$  ( $p = 0,016$ ; zweiseitig). Dies entspricht nach Cohen (1988) einer mittleren Effektstärke [43].

Der Studienfortschritt korrelierte außerdem signifikant positiv mit den Punkten in den Kategorien Bild ( $\rho = 0,34$ ;  $p = 0,034$ ) und System ( $\rho = 0,40$ ;  $p = 0,011$ ). Auch hier ist nach

Cohen (1988) von einer mittleren Effektstärke zu reden [43]. Bei den Punkten in den Kategorien Gerät ( $p = 0,059$ ) und Interpretation ( $p = 0,108$ ) lag keine signifikante Korrelation zum Studienfortschritt vor. Tabelle 13 zeigt die Zusammenfassung der Ergebnisse der Rangkorrelationsanalyse nach Spearman zwischen dem Studienfortschritt und den Punkten.

Tabelle 12: Ergebnisse der Rangkorrelationsanalyse nach Spearman zwischen dem Studienfortschritt und den Punkten

	<b>Korrelation zum Studienfortschritt</b>	<b>Signifikanz (zweiseitig)</b>
<b>Gerät</b>	$\rho = 0,30$	$p = 0,059$
<b>Bild</b>	$\rho = 0,34$	$p = 0,034$
<b>System</b>	$\rho = 0,40$	$p = 0,011$
<b>Interpretation</b>	$\rho = 0,38$	$p = 0,108$
<b>Gesamtpunkte</b>	$\rho = 0,39$	$p = 0,016$

$\rho =$  Korrelationskoeffizient Rho nach Spearman

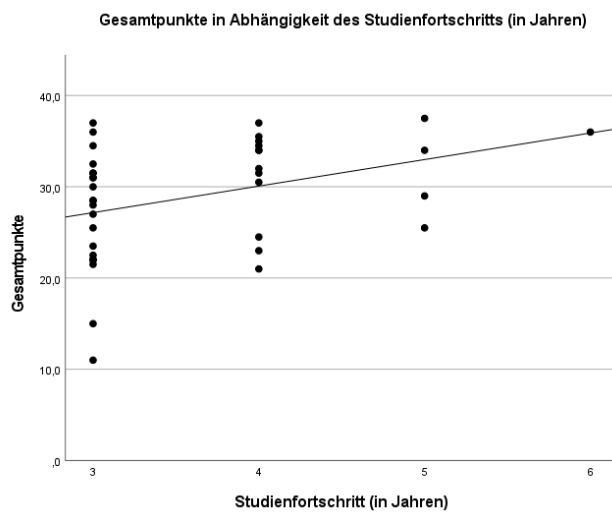


Abbildung 11: Streudiagramm der Gesamtpunkte in Abhängigkeit des Studienfortschritts (in Jahren)

### 3.6 Gruppengröße

Die Proband:innen in den Dreiergruppen ( $n = 9$ ) erreichten im Mittel  $27,5 (\pm 3,5)$  Gesamtpunkte. In den Vierergruppen ( $n = 20$ ) wurden durchschnittlich  $27,9 (\pm 7,4)$  Punkte und in den Fünfergruppen ( $n = 10$ ) durchschnittlich  $32,1 (\pm 5,1)$  Punkte erreicht. Dies entspricht  $68,8\%$ ,  $69,8\%$  und  $80,3\%$  der maximal erreichbaren Punkte. Tabelle 14 schlüsselt die Daten nach Gruppengröße auf und Abbildung 12 stellt diese auf einem Streudiagramm graphisch dar.

Tabelle 13: Gesamtpunkte in Abhängigkeit der Gruppengröße

Gruppengröße	n	Gesamtpunkte (M $\pm$ SD)
3 Personen	9	27,5 ( $\pm$ 3,5) 68,8%
4 Personen	20	27,9 ( $\pm$ 7,4) 69,8%
5 Personen	10	32,1 ( $\pm$ 5,1) 80,3%

*n = Proband:innenanzahl, M =Mittelwert,  
SD = Standardabweichung*

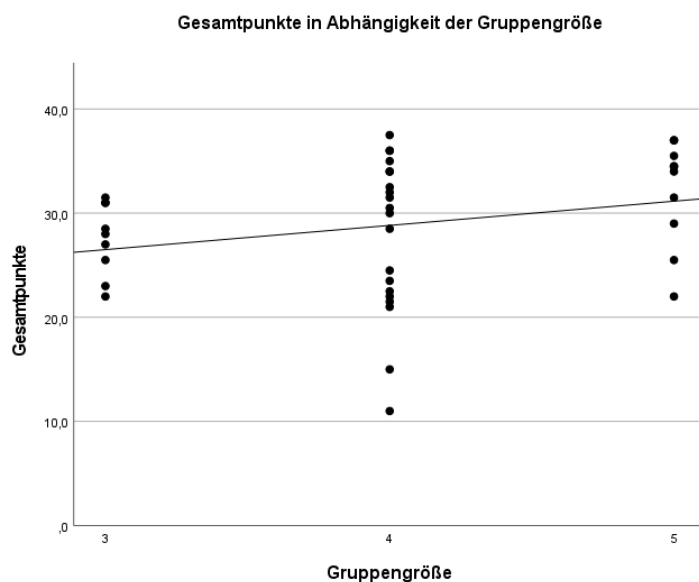


Abbildung 12: Streudiagramm der Gesamtpunkte in Abhängigkeit der Gruppengröße

Da laut eines Shapiro-Wilk-Testes nicht bei allen Gruppengrößen eine Normalverteilung der Punkte angenommen werden konnte, wurde zur Analyse der Abhängigkeit der Gesamtpunkte von der Gruppengröße die Rangkorrelationsanalyse nach Spearman verwendet [42]. Hierbei korrelierten die Gesamtpunkte signifikant positiv mit der Gruppengröße. Der Korrelationskoeffizient Rho nach Spearman betrug  $\rho = 0,33$  ( $p = 0,038$ ; zweiseitig), entsprechend einem mittleren Effekt nach Cohen (1988) [43].

Die Gruppengröße korrelierte außerdem signifikant positiv mit den Punkten in den Kategorien Bild ( $\rho = 0,34$ ;  $p = 0,032$ ) und System ( $\rho = 0,45$ ;  $p = 0,004$ ). Auch hier ist nach Cohen (1988) von einer mittleren Effektstärke zu reden [43]. Bei den Punkten in den Kategorien Gerät ( $p = 0,186$ ) und Interpretation ( $p = 0,242$ ) lag keine signifikante Korrelation zu der Gruppengröße vor. Tabelle 15 zeigt die Zusammenfassung der Ergebnisse der Rangkorrelationsanalyse nach Spearman zwischen der Gruppengröße und den Punkten.

Tabelle 14: Ergebnisse der Rangkorrelationsanalyse nach Spearman zwischen der Gruppengröße und den Punkten

	<b>Korrelation zur Gruppengröße</b>	<b>Signifikanz (zweiseitig)</b>
<b>Gerät</b>	$\rho = 0,22$	$p = 0,186$
<b>Bild</b>	$\rho = 0,34$	$p = 0,032$
<b>System</b>	$\rho = 0,45$	$p = 0,004$
<b>Interpretation</b>	$\rho = 0,19$	$p = 0,242$
<b>Gesamtpunkte</b>	$\rho = 0,33$	$p = 0,038$

$\rho =$  Korrelationskoeffizient Rho nach Spearman

## 3.7 Prüfungszeit

### 3.7.1 Deskriptive Statistik

Die Proband:innen benötigten im Mittel 5,6 ( $\pm 1,7$ ) Minuten zur Absolvierung der gesamten Prüfung. Die minimal benötigte Zeit betrug 2,0 Minuten. Die maximal vorgegebene Zeit von 8,0 Minuten wurde von insgesamt 7 Proband:innen (17,9%) in der Studienpopulation voll ausgeschöpft. Die Proband:innen in der Interventionsgruppe benötigten im Mittel 5,4 ( $\pm 1,8$ ) Minuten, jene in der Kontrollgruppe durchschnittlich 5,6 ( $\pm 1,7$ ). In Tabelle 16 sind die Prüfungszeiten nach Studiengruppe aufgeschlüsselt.

Tabelle 15: Prüfungszeit (in Minuten) in Abhängigkeit der Studiengruppe

	n	Prüfungszeit (M ± SD)	Prüfungszeit (Minimum)	Prüfungszeit (Maximum)
<b>Intervention</b>	20	5,4 (± 1,8)	2,0	8,0
<b>Kontrolle</b>	19	5,6 (± 1,7)	2,5	8,0
<b>Studienpopulation</b>	39	5,6 (± 1,7)	2,0	8,0

*n = Proband:innenanzahl, M = Mittelwert, SD = Standardabweichung*

### 3.7.2 Unterschiede zwischen den Studiengruppen

Da ein Shapiro-Wilk-Test zeigte, dass keine Normalverteilung der Zeiten angenommen werden konnte, wurde zur Analyse der Unterschiede der Prüfungszeit zwischen den Studiengruppen der Mann-Whitney-U-Test gewählt [42]. Dieser zeigte keinen signifikanten Unterschied in der Prüfungszeit zwischen den Studiengruppen ( $U = 181,000$ ;  $z = -0,255$ ;  $p = 0,813$ ; zweiseitig exakt).

### 3.7.3 Zusammenhang zu den Gesamtpunkten

Ein Shapiro-Wilk-Test zeigte, dass die Punkte nicht in allen Kategorien als normalverteilt angenommen werden können. Auch die Prüfungszeit ist nicht als normalverteilt anzunehmen. Deswegen wurde zur Dependenzanalyse die Rangkorrelationsanalyse nach Spearman gewählt [42]. Die erreichten Gesamtpunkte korrelierten hierbei negativ mit der benötigten Prüfungszeit. Der Korrelationskoeffizient Rho nach Spearman betrug  $\rho = -0,64$  ( $p < 0,001$ ; zweiseitig). Dies entspricht einem starken Effekt nach Cohen (1988) [43]. Abbildung 13 zeigt den Zusammenhang zwischen Prüfungszeit und Gesamtpunkten in einem Streudiagramm.

Die Korrelationskoeffizienten zwischen den Punkten in den einzelnen Kategorien und der Prüfungszeit lagen zwischen  $\rho = -0,48$  und  $\rho = -0,70$  ( $p < 0,001$ ; zweiseitig) und können Tabelle 17 entnommen werden. Auch dies entspricht einem starken Effekt nach Cohen (1988) [43].

Tabelle 16: Korrelation nach Spearman zwischen der Prüfungszeit (in Minuten) und den Punkten

	Korrelation zur Prüfungszeit	Signifikanz (zweiseitig)
<b>Gerät</b>	$\rho = -0,61$	$p < 0,001$
<b>Bild</b>	$\rho = -0,70$	$p < 0,001$
<b>System</b>	$\rho = -0,66$	$p < 0,001$
<b>Interpretation</b>	$\rho = -0,48$	$p < 0,001$
<b>Gesamtpunkte</b>	$\rho = -0,64$	$p < 0,001$

$\rho =$  Korrelationskoeffizient Rho nach Spearman

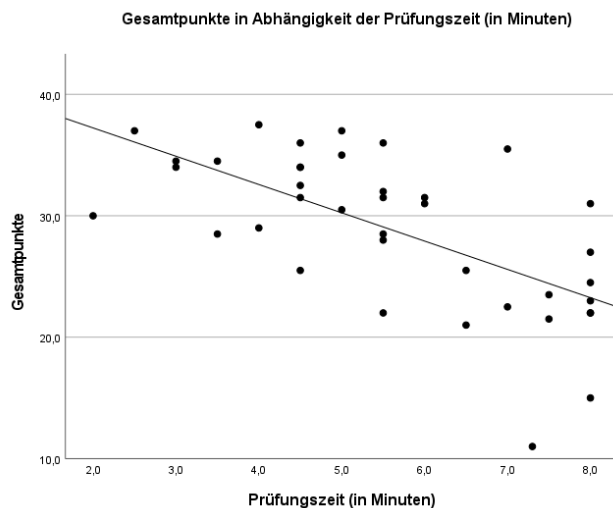


Abbildung 13: Streudiagramm der Gesamtpunkte in Abhängigkeit der Prüfungszeit (in Minuten)

## 3.8 Prüfungsaufgaben

### 3.8.1 Verteilung der Prüfungsaufgaben innerhalb der Studiengruppen

Die insgesamt häufigste Prüfungsaufgabe war die Gallenblase, welche 14 mal (17,9%) in der Studienpopulation absolviert wurde. In der Interventionsgruppe waren die Gallenblase ( $n = 6$ ) die häufigste Aufgabe (15,0%) und die Aorta abdominalis ( $n = 3$ ) die seltenste (7,5%). In der Kontrollgruppe waren die Gallenblase ( $n = 8$ ) die häufigste Aufgabe (21,1%) und die Vena portae ( $n = 3$ ) sowie die rechte Niere ( $n = 3$ ) die seltensten Aufgaben (je 7,9%). Die Häufigkeiten der gezogenen Prüfungsaufgaben in den beiden Studiengruppen sowie in der gesamten Studienpopulation können Tabelle 18 entnommen werden.

Zum Ausschluss, dass die verschiedenen Prüfungsaufgaben ungleichmäßig auf die beiden Studiengruppen verteilt waren, wurde eine Kontingenzanalyse durchgeführt. Aufgrund der Stichprobengröße ( $n = 78$ ), den Freiheitsgraden ( $df = 7$ ) und der Erwartungswerte (12 Zellen mit einer erwarteten Häufigkeit  $< 5$ ) wurde der exakte Test nach Fisher gewählt [42]. Hier zeigten sich keine signifikanten Unterschiede in der Verteilung der Prüfungsaufgaben auf die beiden Studiengruppen ( $p = 0,746$ ; zweiseitig).

Tabelle 17: Häufigkeiten der Prüfungsaufgaben

<b>Prüfungsaufgabe</b>		<b>Intervention (n = 40)</b>	<b>Kontrolle (n = 38)</b>	<b>Studienpopulation (n = 78)</b>
<b>Gallenblase</b>	Anzahl	6	8	14
	Prozent	15,0%	21,1%	17,9%
<b>Vena cava inferior</b>	Anzahl	5	4	9
	Prozent	12,5%	10,5%	11,5%
<b>Rechte Niere</b>	Anzahl	5	3	8
	Prozent	12,5%	7,9%	10,3%
<b>Linke Niere</b>	Anzahl	6	2	8
	Prozent	15,0%	5,3%	10,3%
<b>Lebervenen</b>	Anzahl	6	7	13
	Prozent	15,0%	18,4%	16,7%
<b>Vena portae</b>	Anzahl	5	3	8
	Prozent	12,5%	7,9%	10,3%
<b>Milz</b>	Anzahl	4	5	9
	Prozent	10,0%	13,2%	11,5%
<b>Aorta abdominalis</b>	Anzahl	3	6	9
	Prozent	7,5%	15,8%	11,5%
<b>Summe</b>	Anzahl	40	38	78
	Prozent	100%	100%	100%

*n = Anzahl an absolvierten Prüfungsaufgaben*

### 3.8.2 Zeit in Abhängigkeit der Prüfungsaufgabe

Die 39 Proband:innen absolvierten je 2 Aufgaben, in Summe also 78 Aufgaben. Im Durchschnitt wurden 2,8 ( $\pm 1,2$ ) Minuten pro Aufgabe benötigt. Die kürzeste benötigte Zeit war 0,5 Minuten für die Darstellung der Vena portae. Die Aufgabe betreffend der Vena portae hatte auch die niedrigste maximal benötigte Zeit von 3,5 Minuten. Die höchste maximal benötigte Zeit waren 6,0 Minuten für die Darstellung der Lebervenen und der Aorta abdominalis.

Die Darstellung der linken Niere stach besonders ins Auge. Hier benötigten die Proband:innen im Mittel 4,0 ( $\pm 0,7$ ) Minuten. Die minimal benötigte Zeit lag bei 3,0 Minuten, während bei allen anderen Aufgaben die minimale Zeit bei 0,5 bis 2,0 Minuten lag. Tabelle 19 und Abbildung 14 präsentieren tabellarisch respektive als Boxplots die Zeit in Abhängigkeit der Prüfungsaufgabe.

Da ein Shapiro-Wilk-Test zeigte, dass die Zeit nicht bei allen Prüfungsaufgaben normalverteilt war, wurde aufgrund der Stichprobengröße ( $n = 78$ ) und den Freiheitsgraden ( $df = 7$ ) der Kruskal-Wallis-H-Test zur Dependenzanalyse gewählt. Da mehr als 5 Gruppen vorlagen, wurde die Teststatistik als asymptotisch Chi-Quadrat verteilt angenommen [42]. Der Kruskal-Wallis-H-Test zeigte signifikante Unterschiede bezüglich der Mittelwerte der benötigten Zeit in Abhängigkeit der Prüfungsaufgabe ( $\text{Chi}^2 = 21,679$ ;  $p = 0,003$ ; asymptotisch zweiseitig).

Im Post-hoc-Test nach Dunn-Bonferroni zeigten sich signifikante Unterschiede in der benötigten Zeit lediglich zwischen der Darstellung der Vena cava inferior und der linken Niere ( $n = 17$ ;  $z = -3,681$ ;  $p = 0,007$ ; zweiseitig) sowie zwischen der Vena portae und der linken Niere ( $n = 16$ ;  $z = 3,578$ ;  $p = 0,010$ ; zweiseitig).

Aus den Standardteststatistiken des Dunn-Bonferroni-Tests wurden Korrelationskoeffizienten nach Pearson berechnet [42]. Es handelt sich sowohl zwischen den Aufgaben „Vena cava inferior“ und „Linke Niere“ ( $r = 0,89$ ) als auch zwischen „Vena portae“ und „Linker Niere“ ( $r = 0,89$ ) um einen starken Effekt nach Cohen (1988) [43].

Tabelle 18: Zeit pro Aufgabe (in Minuten) in Abhängigkeit der Prüfungsaufgabe (M = Mittelwert, SD = Standardabweichung)

Prüfungsaufgabe	n	Zeit (M ± SD)	Zeit (Minimum)	Zeit (Maximum)
Gallenblase	14	3,2 (± 1,1)	1,5	5,0
Vena cava inferior	9	1,9 (± 1,1)	1,0	4,0
Rechte Niere	8	2,8 (± 1,1)	1,0	5,0
Linke Niere	8	4,0 (± 0,7)	3,0	5,5
Lebervenen	13	2,6 (± 1,3)	1,0	6,0
Vena portae	8	1,9 (± 0,9)	0,5	3,5
Milz	9	2,6 (± 1,1)	1,0	4,0
Aorta abdominalis	9	3,2 (± 1,2)	2,0	6,0
Studienpopulation	78	2,8 (± 1,2)	0,5	6,0

*n = Anzahl an absolvierten Prüfungsaufgaben*

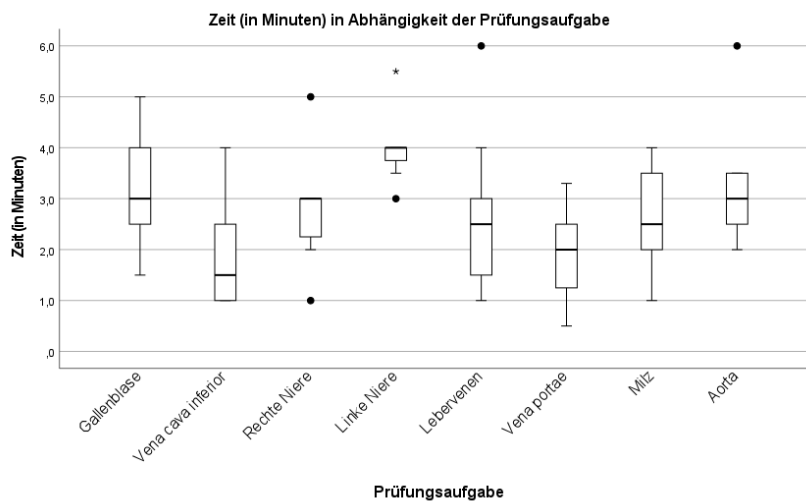


Abbildung 14: Boxplot der Zeit pro Aufgabe (in Minuten) in Abhängigkeit der Prüfungsaufgabe

### 3.9 Prüfer:innen

Zur Beurteilung der Zuverlässigkeit des Bewertungsschemas wurden jeweils die Punkte, die jede:r Proband:in bei den beiden Prüfer:innen erreichte, miteinander verglichen. Da jedes Video nur von je zwei der insgesamt vier Prüfer:innen bewertet wurde, wurden die Punkte dieser beiden Prüfer:innen miteinander korreliert. Da diese beiden rechnerisch erzeugten

Prüfer:innen nicht einzelnen realen Personen entsprechen, wird hierzu keine gesonderte deskriptive Statistik ausgewiesen.

Im Shapiro-Wilk-Test zeigte sich, dass die Punkte nicht in allen Kategorien als normalverteilt angenommen werden können. Aus diesem Grund wurde zur Dependenzanalyse die Rangkorrelationsanalyse nach Spearman zwischen den vergebenen Punkten der beiden Prüfer:innen pro Proband:in berechnet [42]. Die Gesamtpunkte der beiden Prüfer:innen korrelierten mit einem Korrelationskoeffizienten Rho nach Spearman von  $\rho = 0,70$  ( $p < 0,001$ ). Auch die Punkte in den Kategorien Bild, System und Interpretation korrelierten hochpositiv zwischen den beiden Prüfer:innen mit Korrelationskoeffizienten jeweils zwischen  $\rho = 0,60$  und  $\rho = 0,77$  ( $p < 0,001$ ). Einzig in der Kategorie Gerät war die Korrelation deutlich niedriger mit  $\rho = 0,35$  ( $p = 0,028$ ). Tabelle 20 zeigt eine Übersicht der Ergebnisse der Rangkorrelationsanalyse. Die Abbildungen 15 und 16 stellen die Zusammenhänge zwischen den beiden Prüfer:innen graphisch in Streudiagrammen dar.

*Tabelle 19: Ergebnisse der Rangkorrelationsanalyse nach Spearman zwischen den Punkten der beiden Prüfer:innen je Proband:in*

	<b>Korrelation zwischen den Prüfer:innen</b>	<b>Signifikanz (zweiseitig)</b>
<b>Gerät</b>	$\rho = 0,35$	$p = 0,028$
<b>Bild</b>	$\rho = 0,68$	$p < 0,001$
<b>System</b>	$\rho = 0,77$	$p < 0,001$
<b>Interpretation</b>	$\rho = 0,60$	$p < 0,001$
<b>Gesamtpunkte</b>	$\rho = 0,70$	$p < 0,001$

*$\rho =$  Korrelationskoeffizient Rho nach Spearman*

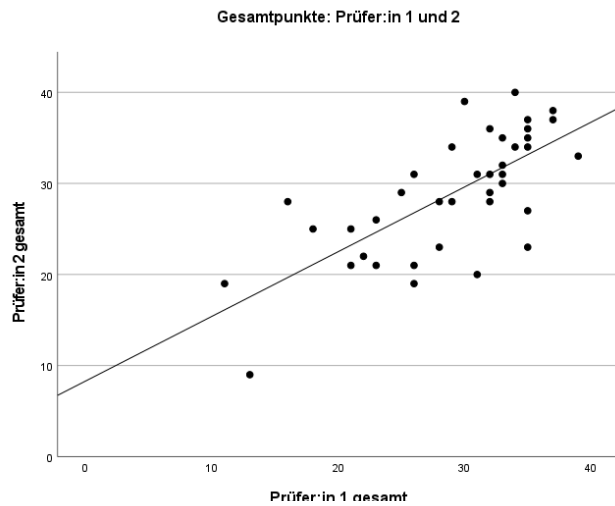


Abbildung 15: Streudiagramm der Gesamtpunkte bei den beiden Prüfer:innen

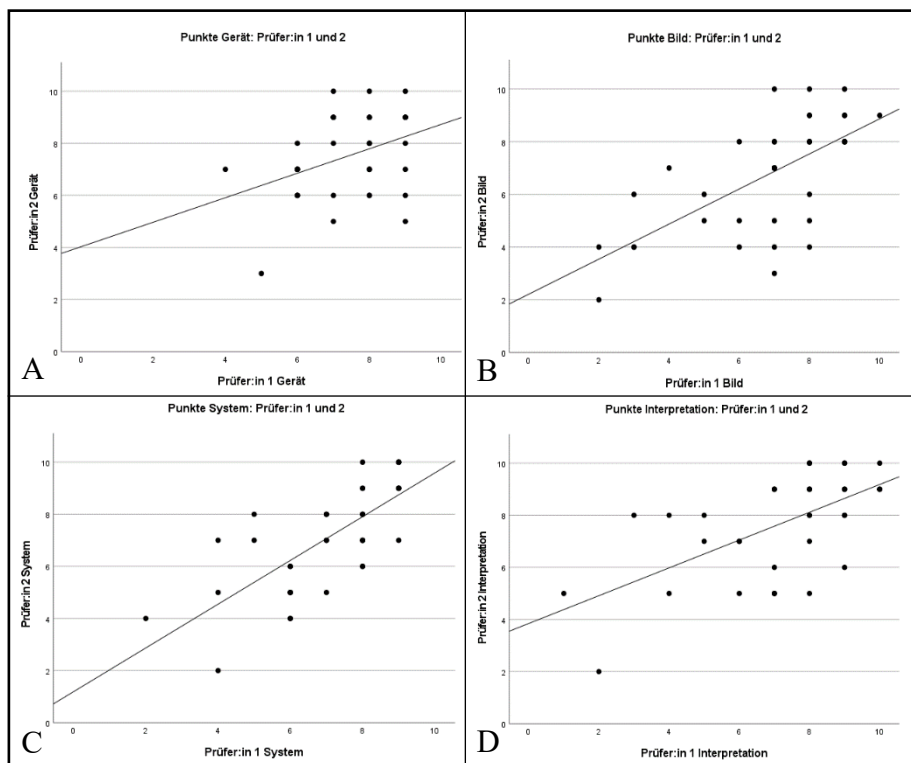


Abbildung 16: Streudiagramme der Punkte zwischen den beiden Prüfer:innen in den Kategorien (A) Gerät, (B) Bild, (C) System, (D) Interpretation

### 3.10 Kategorien

Zur Beurteilung, wie sich die Punkte einer jeden Probandin und eines jeden Probanden in den Kategorien zueinander verhielten, wurden diese jeweils miteinander und mit den Gesamtpunkten korreliert.

Ein Shapiro-Wilk-Test zeigte, dass die Punkte nicht in allen Kategorien normalverteilt waren. Deswegen wurde die Rangkorrelationsanalyse nach Spearman verwendet [42]. Die Punkte in den Kategorien korrelierten untereinander hoch positiv mit Korrelationskoeffizienten Rho nach Spearman zwischen  $\rho = 0,78$  und  $\rho = 0,88$  (jeweils  $p < 0,001$ ). Tabelle 21 zeigt eine Übersicht der Ergebnisse der Rangkorrelationsanalyse.

Tabelle 20: Ergebnisse der Rangkorrelationsanalyse nach Spearman zwischen den Punkten der jeweiligen Kategorien in der Studienpopulation.

	<b>Gerät</b>	<b>Bild</b>	<b>System</b>	<b>Interpretation</b>
<b>Gerät</b>	$\rho = 1$			
<b>Bild</b>	$\rho = 0,81$	$\rho = 1$		
<b>System</b>	$\rho = 0,88$	$\rho = 0,87$	$\rho = 1$	
<b>Interpretation</b>	$\rho = 0,83$	$\rho = 0,78$	$\rho = 0,80$	$\rho = 1$

$\rho =$  Korrelationskoeffizient Rho nach Spearman

## **4. Diskussion**

### **4.1 Ergebnisse**

#### **4.1.1 Fallzahl und Drop-out-Rate**

Die in der Poweranalyse berechnete Fallzahl von 23 Proband:innen pro Gruppe konnte nicht zu Gänze erreicht werden. Dies lag vordergründig an der Drop-out-Rate von 18,8% in der Studienpopulation, welche deutlich über der a priori Schätzung von 5,0% lag. Positiv zu erwähnen ist, dass alle Prüfungsvideos für die Beurteilung verwertbar waren. In mehreren Arbeiten hatte es aufgrund technischer Probleme noch ein Drop-out nach den Prüfungen gegeben [28,29,31].

Mögliche Gründe für die hohe Drop-out-Rate könnte der weniger verbindliche Charakter der Kurse im Vergleich zu den offiziellen Kursen von Sono4you Graz gewesen sein. Letztere sind überaus begehrt und die begrenzten Plätze werden verlost, wodurch sich häufig eine Wartezeit ergibt und die Kurse womöglich deswegen seltener abgebrochen werden. Ein weiterer Grund könnte der Studienzeitpunkt im Juni gewesen sein. Das warme Sommerwetter und die nahende Prüfungsphase könnte sich negativ auf die Adhärenz der Proband:innen ausgewirkt haben. Auch die Rekrutierung gestaltete sich schwieriger als ursprünglich angenommen. Eine Verlängerung des Rekrutierungszeitraums war aufgrund des straffen Zeitplans nicht möglich. Eine weitere Überlegung war, dass gegebenenfalls auch die mobilen Ultraschallgeräte mitverantwortlich für die hohe Drop-out-Rate waren. Diese wurden von den Proband:innen subjektiv als minderwertig eingestuft. Deren Limitationen werden in Abschnitt 4.2.1.2 besprochen.

Die Bedenken, dass es zu einem signifikant höherem Drop-out in der Interventionsgruppe aufgrund der geringeren Attraktivität des Selbststudiums im Vergleich zur Kontrollgruppe kommen könnte, bestätigten sich nicht. Ein exakter Test nach Fisher zeigte in Abschnitt 3.1 keinen signifikanten Unterschied in der Drop-out-Rate zwischen den Studiengruppen ( $p = 0,466$ ).

#### **4.1.2 Demographische Daten**

##### **4.1.2.1 Geschlecht**

Die Kontinuitätskorrektur nach Yates in Abschnitt 3.2.1 ergab eine signifikante Ungleichverteilung der Geschlechter auf die beiden Studiengruppen mit einer weiblichen

Mehrheit in der Kontrollgruppe ( $p = 0,027$ ). Dieses Ungleichgewicht war möglich, da bei der Randomisierung nicht nach Geschlecht stratifiziert wurde. Diese Limitation der Studie ist jedoch in Hinblick auf den Hauptzielparameter nur als gering zu bewerten, da ein T-Test für unabhängige Stichproben in Abschnitt 3.4 keinen signifikanten Unterschied in den Mittelwerten der Gesamtpunkte zwischen den beiden Geschlechtern zeigte ( $p = 0,263$ ).

#### **4.1.2.2 Studienfortschritt**

Es waren hauptsächlich Studierende aus dem 3. und 4. Studienjahr und einzelne Proband:innen aus dem 5. und 6. Studienjahr vertreten. Die Rekrutierung einer homogeneren Studienpopulation bezüglich des Studienfortschritts war nicht möglich.

Die Rangkorrelationsanalyse nach Spearman in Abschnitt 3.5 ergab eine signifikante Korrelation zwischen dem Studienfortschritt und den Gesamtpunkten ( $\rho = 0,39$ ;  $p = 0,016$ ). Dies mindert jedoch nicht die Aussagekraft des Hauptzielparameter, da ein exakter Test nach Fisher in Abschnitt 3.2.2 zeigte, dass es keine signifikante Ungleichverteilung des Studienfortschritts auf die beiden Studiengruppen gab ( $p = 0,891$ ).

Die Korrelation zwischen Studienfortschritt und Gesamtpunkten lässt sich in erster Linie auf die verschiedenen Vorerfahrungen der Proband:innen zurückführen. Alle Proband:innen hatten bereits vor der Studie zumindest einführenden Kontakt zur Abdomensonographie im Rahmen der curricularen Lehre an der Medizinischen Universität Graz gehabt. Im Laufe des fortschreitenden Studiums akkumulierten die Studierenden durch Seminare und Praktika immer mehr Wissen bezüglich der Abdomensonographie, auch wenn sie keine spezifischen Kurse dazu absolvierten. So ist es plausibel, dass die steigende Exposition zu Ultraschall mit steigendem Studienfortschritt die Gesamtpunkte positiv beeinflusste. In diesem Sinn zeigte sich die Studie als „Progress Test“, da bei gleichen Testanforderungen die Testergebnisse mit den Studienjahren anstiegen. Diese Fragestellung war zwar keines der Ziele der Studie, zeigt jedoch, dass Studierende der Humanmedizin an der Medizinischen Universität Graz in der curricularen Lehre nach und nach praktisch anwendbare Kenntnisse und Fertigkeiten in der Abdomensonographie aufbauen.

Zusätzlich zum vermutlich größeren Vorwissen der fortgeschrittenen Proband:innen zeigte Murad et al. (2010) in einem systematischen Review, dass ein höherer Studienfortschritt mit höherer Bereitschaft und Effektivität für das selbstständige Erarbeiten von Lehrinhalten einhergeht [10].

### **4.1.2.3 Gruppengröße**

In der Rangkorrelationsanalyse nach Spearman in Abschnitt 3.6 korrelierten die Gesamtpunkte signifikant positiv mit der Gruppengröße ( $\rho = 0,333$ ;  $p = 0,038$ ). Ein Erklärungsansatz dafür ist, dass womöglich in zu kleinen Gruppen weniger Dynamik herrschte und sich die Proband:innen so weniger gegenseitig motivieren und weiterhelfen konnten.

Noerholk et al. (2022) fand in einer Untersuchung keine Unterschiede in den Gesamtpunkten zwischen Zweier-, Dreier- und Vierergruppen. Es wurde beobachtet, dass die kürzere Hands-on-Zeit in den größeren Gruppen durch konstruktive und interaktive Lern-Aktivitäten kompensiert wurden [30]. Ähnliche Dynamiken lagen vermutlich auch in dieser Studie vor, welche zu einer Überlegenheit der größeren Gruppen führte. Die positive Korrelation lässt sich vermutlich nur bedingt nach oben extrapolieren. Es ist anzunehmen, dass ab einer gewissen Gruppengröße der Lernerfolg wieder sinkt, da die verkürzte individuelle Übungszeit dann die gruppenspezifischen Vorteile aufwiegt.

Die Aussagekraft der Korrelation zwischen Gesamtpunkten und Gruppengröße bei dieser Studie ist durch die mangelnde Konstanz der Gruppengröße im Studienverlauf deutlich limitiert. Durch die hohe Drop-out-Rate verkleinerten sich einige Gruppen im Laufe der Studie.

### **4.1.2 Hauptergebnisse**

Es bestand die Hypothese, dass die Kontrollgruppe, welche in intensiver persönlicher Interaktion mit dem Tutor stand, mit einer Effektstärke von  $d = 1,0$  besser abschneiden würde als die Interventionsgruppe, die sich überwiegend auf Selbststudium mit dem E-Learning stützte. Diese Behauptung kann ganz klar verworfen werden, da sich bei einseitiger Testung keine signifikante Überlegenheit der Kontrollgruppe im Mann-Whitney-U-Test zeigte ( $p = 0,061$ ). Die fehlende Überlegenheit der Kontrollgruppe wird vorrangig mit dem sehr kontrollierten Studienprotokoll der Interventionsgruppe erklärt. Die diesbezüglichen Limitationen werden in Abschnitt 4.2.1.3 erläutert.

Die Gesamtpunkte aus dieser Arbeit können nicht ohne weiteres mit jenen aus anderen Arbeiten verglichen werden, welche ebenfalls mit dem Beurteilungsschema OSAUS arbeiteten, da die Punktevergabe jeweils verschieden gehandhabt wurde. Aufgrund wechselnder Anzahl inkludierter Kategorien sind die absoluten Werte nicht vergleichbar. Auch beim Vergleich der relativen Werte muss bedacht werden, dass in dieser Studie die

Vergabe von null Punkten möglich war, während in den Vergleichsarbeiten mindestens ein Punkt vergeben werden musste.

### **4.1.3 Nebenergebnisse**

#### **4.1.3.1 Punkte in der Kategorie Gerät**

Die Punkte der beiden Prüfer:innen je Proband:in korrelierten in der Kategorie Gerät ( $\rho = 0,35$ ;  $p = 0,028$ ), in welcher der Umgang mit dem Gerät, den Knöpfen und dem Schallkopf bewertet wurde, deutlich niedriger miteinander, als in den übrigen Kategorien (zwischen  $\rho = 0,60$  und  $\rho = 0,77$ ; jeweils  $p < 0,001$ ).

Ein Erklärungsansatz dafür ist, dass die Prüfer:innen selbst vorrangig an traditionellen Ultraschallgeräten lehrend tätig waren und somit weniger vertraut mit den Details und Limitationen der mobilen Butterfly IQ+ Sonden waren. Außerdem kam von den Prüfer:innen die Rückmeldung, dass die Kameraposition das Beurteilen des Umgangs mit dem Gerät erschwerte. Auch die adäquate Menge an Schallgel und Druck auf den Schallkopf wäre durch die Videodokumentation schwierig zu beurteilen gewesen.

#### **4.1.3.2 Korrelation der Punkte in den Kategorien**

Die Punkte in den einzelnen Kategorien korrelierten in der Rangkorrelationsanalyse nach Spearman untereinander hoch positiv miteinander (zwischen  $\rho = 0,78$  und  $\rho = 0,88$ ; jeweils  $p < 0,001$ ). Der führende Erklärungsansatz für diese hohen Korrelationskoeffizienten ist, dass die Proband:innen in allen vier Kategorien jeweils gleich kompetent oder gleich inkompetent waren. Es ist beispielsweise anzunehmen, dass bei mangelnder Gerätekunde keine adäquate Bildoptimierung möglich ist. Ebenso wird eine wenig systematische Untersuchung seltener in einem zufriedenstellenden Bild resultieren. Ist das Bild in seiner Aussagekraft eingeschränkt, so ist auch eine ausreichende Interpretation nur erschwert möglich.

Ein gewisser Halo-Effekt vonseiten der Prüfer:innen kann jedoch nicht ausgeschlossen werden. Beim Halo-Effekt handelt es sich um eine Verzerrung, bei der von einem Merkmal ausgehend ohne objektive Grundlage auf weitere Eigenschaften geschlossen wird. Dabei orientieren sich Personen am Gesamteindruck, anstatt die Merkmale getrennt zu beurteilen [44]. Sehr gut erforscht ist dieser Effekt beim Merkmal der physischen Attraktivität. Nachdem die Prüfer:innen die Proband:innen nicht persönlich getroffen haben, sondern die Beurteilung lediglich anhand der Videoaufzeichnungen erfolgt ist, auf denen überdies die

Gesichter der Proband:innen nicht zu sehen waren, ist die Verzerrung durch die Attraktivität der Proband:innen in dieser Studie als äußerst gering einzuschätzen.

Die Verzerrung durch einen möglichen Halo-Effekt wurde in keiner der bisher veröffentlichten Arbeiten zum Beurteilungsschema OSAUS berücksichtigt. Für zukünftige Studien würde sich das zeitlich getrennte Beurteilen der einzelnen Kategorien nacheinander empfehlen, oder auch das Beurteilen jeder Kategorie durch jeweils andere Prüfer:innen, um das Risiko einer Verzerrung zu minimieren [44].

### **4.1.3.3 Prüfungszeit**

Es zeigte sich eine negative Korrelation der Prüfungszeit mit den Gesamtpunkten ( $\rho = -0,64$ ;  $p < 0,001$ ). Dies bestätigt die Hypothese, dass kompetentere Proband:innen weniger Zeit für die Absolvierung der Aufgaben benötigen, da sie sowohl im Auffinden der gesuchten Struktur als auch in der Bildoptimierung sowie der Interpretation schneller als weniger kompetente Proband:innen sind.

Der Beurteilung der Prüfungszeit sollte jedoch nicht allzu viel Gewicht beigemessen werden, da sie einigen Limitationen unterliegt. Aufgrund der geringen Trennschärfe bezüglich des genauen Start- und Endzeitpunkts einer Prüfungsaufgabe wurde die Zeit auf die nächste halbe Minute aufgerundet dokumentiert. Eine exaktere Zeitnehmung wäre nicht seriös durchführbar gewesen. Diese Unschärfe ist aber insofern zu vernachlässigen, als die benötigte Zeit nicht als beurteilungsrelevant definiert wurde und lediglich der Vollständigkeit halber dokumentiert wurde. Da dies den Proband:innen auch vorab so mitgeteilt wurde, verbrauchten einige Proband:innen ihre gesamte restliche Zeit bei der zweiten Aufgabe, um ein bereits akzeptables Bild, welches sie unter anderen Umständen bereits dokumentiert hätten, noch weiter zu verbessern.

## **4.2 Limitationen**

### **4.2.1 Methoden**

#### **4.2.1.1 E-Learning**

Eine der Limitationen dieser Studie ist, dass das verwendete E-Learning auf ein Flipped-Classroom-Format und nicht auf das Selbststudium zugeschnitten war. Aus den Rückmeldungen der Proband:innen der Interventionsgruppe ging hervor, dass diese sich unter anderem die Inklusion von Videos gewünscht hätten, in denen die wichtigsten Fehler

und deren Adressierung demonstriert werden. Dies spiegelt auch die Eindrücke des Autors wider, dass die Proband:innen in der Interventionsgruppe Probleme hatten, Fehler zu identifizieren und diese zu korrigieren. Außerdem wurden neben Beispielbildern unter optimalen Bedingungen auch solche unter schwierigen, realistischeren Schallbedingungen gefordert.

In folgenden Arbeiten zum Selbststudium in der Sonographie sollten eigens darauf zugeschnittene Lernunterlagen verwendet werden. Murad et al. (2010) zeigten zudem in einem systematischen Review, dass das Selbststudium in der medizinischen Didaktik effektiver war, wenn die Lernenden in die Auswahl der Lernunterlagen miteinbezogen wurden [10].

#### **4.2.1.2 Ultraschallgeräte**

Eine weitere Limitation der Studie bestand in den verwendeten mobilen Ultraschallgeräten. Einige Proband:innen vermeldeten, dass sie im Vergleich zu traditionellen Geräten durch die subjektiv empfundene niedrigere Qualität einen niedrigeren Lernerfolg hatten. Dewar et al. (2020) hatten im Vergleich zwischen dem Butterfly IQ, dem Vorgängermodell des Butterfly IQ+, und einem traditionellen Mittelklasse-Ultraschallgerät keinen signifikanten Unterschied in der Bildqualität und Akquirierungszeit gefunden [45]. Jedoch wurde jene Studie nicht mit Lernenden, sondern mit einem einzigen Arzt in der Ausbildung zum Notfallmediziner durchgeführt. Die von den Proband:innen subjektiv empfundene niedrigere Qualität lag somit gegebenenfalls am Ausgabegerät, welches einen etwas kleineren Bildschirm als ein handelsübliches Ultraschallgerät besaß.

Es wäre eine interessante Fragestellung, ob sich die Art der verwendeten Ultraschallgeräte auf den Lernerfolg der Proband:innen auswirkt. In dieser Arbeit wurde die Annahme getroffen, dass diesbezüglich kein Unterschied zwischen mobilen Ultraschallgeräten und traditionellen Geräten besteht. Ob die erlernten Fähigkeiten ausreichend auf ein traditionelles Ultraschallgerät transferiert werden können, ist in Zukunft ebenfalls noch zu untersuchen.

#### **4.2.1.3 Interventionsgruppe**

Es wurde in dieser Studie bei der Interventionsgruppe mit dem Selbststudium ein kontrolliertes Setting gewählt, welches in Abschnitt 2.3.2 erläutert wurde. Die Adhärenz der Proband:innen in der Interventionsgruppe wurde bei einem unregulierten, dezentralen, eigenständig organisierten Selbststudium als zu niedrig eingeschätzt. Es wurde bezweifelt,

dass eine zufällig zusammengesetzte Gruppe an Einzelpersonen die räumliche und zeitliche Organisation in einem zur Kontrollgruppe vergleichbaren Rahmen schaffen würde. Als Referenz diente eine Studie, welche aufgrund der hohen Drop-out-Rate abgebrochen werden musste. Dort meldeten die Proband:innen vor allem das mangelnde Feedback und den unterschätzten Zeitaufwand als Ausstiegsgrund retour [14].

Das gewählte Format erfüllte nicht alle drei Hauptkriterien des „Self-directed Learnings“ nach Knowles (1975), welche in Abschnitt 1.2.2 erläutert wurden [8]. Es lässt sich die Frage stellen, inwiefern das kontrollierte Setting den Lernerfolg der Interventionsgruppe beeinflusste. Es ist anzunehmen, dass die Präsenz des Tutors vor Ort einen fördernden Einfluss auf die Motivation der Proband:innen hatte, auch wenn dieser nicht aktiv in den Lernprozess eingriff. Auch die offiziellen Räumlichkeiten der Medizinischen Universität Graz hatten eventuell einen Einfluss auf die Leistungsbereitschaft der Proband:innen.

Es wurde jedoch mit der in dieser Studie gewählten Variante das Grundproblem, dass zu wenig tutorierendes Personal in Relation zur Nachfrage nach Ultraschallkursen zur Verfügung steht, adressiert. Das Studiendesign der Interventionsgruppe lässt sich mit einem geringeren Betreuungsschlüssel als jenes der Kontrollgruppe nach oben skalieren. Letztlich wird die Anzahl der Geräte zum limitierenden Faktor werden.

## **4.2.2 Prüfung des Lernerfolgs**

### **4.2.2.1 Prüfer:innen**

Das Ziel der Studie war, die grundlegenden praktischen Fähigkeiten der Abdomensonographie abzutesten. Da nicht das klinische Wissen im Vordergrund stand, wurde es als akzeptabel angesehen, erfahrene Studierende anstatt Fachärzt:innen für Radiologie als Prüfer:innen einzusetzen. Alle Prüfer:innen brachten als langjährige Tutor:innen bei Sono4you Graz und als zertifizierte studentische Tutor:innen der ÖGUM im Bereich der Abdomensonographie die nötige Expertise zur Beurteilung der Proband:innen mit [46].

Jedes Video wurde von zwei Prüfer:innen bewertet, da Todsén et al. (2015) zeigte, dass bei zwei Prüfer:innen und zwei Prüfungsaufgaben der Generalisationskoeffizient des OSAUS-Bewertungsschemas bei 0,70 liegt, was im Rahmen dieser Studie als ausreichend hoch eingestuft wurde [28]. Auch Bahner et al. (2011) machten bei ihrem Bewertungsschema

„B-QUIET“ die Beobachtung, dass die Reliabilität mit der Einführung einer zweiten prüfenden Person deutlich anstieg [47].

Um eine Verzerrung zu verhindern waren die Prüfer:innen weder in die Betreuung der Proband:innen noch in die Konzeption des Studiendesigns involviert. Zur Verhinderung eines potenziellen Beobachterdrifts wurden je zwei Prüfer:innen angehalten, die Videos mit aufsteigender respektive absteigender Identifikationsnummer zu bearbeiten. Da nicht alle Proband:innen von allen Prüfer:innen bewertet wurden, war die Analyse der einzelnen Prüfer:innen und somit die Berechnung einer Inter-Rater-Reliability nicht möglich.

#### **4.2.2.2 Prüfungsaufgaben**

Es war nicht möglich, alle Proband:innen dieselben Prüfungsaufgaben durchführen zu lassen, da die Gruppenkolleg:innen bei der Prüfung im Raum anwesend waren. Dementsprechend wurde bei der Konstruktion der Prüfungsaufgaben darauf geachtet, diese in einer ähnlichen Schwierigkeitsstufe zu halten. Die Frage, zu welchem Ausmaß dies gelungen ist, konnte nicht abschließend geklärt werden, da die Prüfer:innen die Punkte nicht gesondert pro Prüfungsaufgabe, sondern gemittelt über die gesamte Prüfungssituation vergaben. Aus Abschnitt 3.7.3 ging jedoch hervor, dass die Gesamtpunkte negativ mit der Prüfungszeit korrelierten. Die Hypothese lautete, dass kein signifikanter Unterschied in der benötigten Zeit zwischen den Aufgaben besteht. Diese Hypothese musste in Abschnitt 3.8.2 verworfen werden.

Es ist anzunehmen, dass die Proband:innen jene Aufgaben, für die im Durchschnitt weniger Zeit benötigt wurde, als leichter empfanden. Die Aufgabe „Vena portae“ hatte die kürzeste minimale und kürzeste maximale Prüfungszeit. Beide Fakten sprechen dafür, dass diese die leichteste aller Aufgaben war. Die Aufgabe „Linken Niere“ hatte die höchste mittlere sowie die höchste maximale Prüfungszeit. Entsprechend wurde diese Prüfungsaufgabe die schwierigste interpretiert. Aus den oben genannten Gründen können diese Behauptungen jedoch nicht restlos objektiviert werden.

Dass die Prüfungsaufgaben keine einheitliche Schwierigkeit aufwiesen mindert jedoch die Aussagekraft der Studie bezüglich des Hauptzielparameters nicht schwerwiegend, da der exakte Test nach Fisher in Abschnitt 3.8.1 zeigte, dass keine signifikante Ungleichverteilung der Prüfungsaufgaben auf die beiden Studiengruppen vorlag ( $p = 0,746$ ). Es kann jedoch nicht ausgeschlossen werden, dass tendenziell weniger kompetente Proband:innen zufällig schwierigere Aufgaben erhielten.

Bei zukünftigen Studien, bei denen nicht alle Proband:innen dieselben Prüfungsaufgaben gestellt bekommen, ist noch genaueres Augenmerk darauf zu legen, dass die Aufgaben bezüglich der Schwierigkeit ausgeglichen sind. Alternativ dazu könnten Prüfungsaufgaben je nach Schwierigkeit zu verschiedenen Pools mit gegebenenfalls adaptierten Zeitlimits zusammengefasst werden.

#### **4.2.2.3 Kategorie Interpretation**

Bei der Beurteilung der Punkte in der Kategorie Interpretation gibt es einige Limitationen zu beachten. Es wurden ausschließlich junge, gesunde Proband:innen ohne Pathologien untersucht, wodurch sich diese Kategorie auf Theoriefragen beschränkte.

In der Originalarbeit zum Bewertungsschema OSAUS von Tolsgaard et al. (2013) wurde zur Beurteilung der Kategorie Interpretation verlangt, am Prüfungsende einen Untersuchungsbefund zu diktieren, welcher auf Vollständigkeit, Präzision und Fachjargon beurteilt wurde [23]. In dieser Studie wurde gegen die Verwendung dieser Vorgehensweise entschieden, da das Formulieren radiologischer Befunde nicht ausreichend in den Lernunterlagen behandelt wurde.

Eine weitere mögliche Alternative ist die Interpretation separater Video-Clips, welche physiologische oder pathologische Befunde zeigen, wie sie beispielsweise bei Nilsson et al. (2017) verwendet wurden [29]. Für zukünftige Studien mit ebenfalls ausschließlich gesunden Schallmodellen ist das eine Option, die in Betracht gezogen werden sollte.

#### **4.2.2.4 Schallmodelle**

Bei den Prüfungen wurden ausschließlich junge, gesunde, normalgewichtige Erwachsene ohne relevante Pathologien untersucht. In einige Vorarbeiten wurde an Patient:innen mit fassbaren Pathologien geprüft. Dies ermöglichte auch die Erhebung der diagnostischen Präzision der Proband:innen sowie die Korrelation der Präzision mit den Gesamtpunkten [28,31]. Da jedoch auch in der Kurssituation lediglich an jungen, gesunden Personen geübt wurde und der Fokus der Lehreinheiten auf dem Erwerb grundlegender praktischer Fähigkeiten lag, wurde dieses Vorgehen als zulässig erachtet und steht auch in Konkordanz zu anderen Studien, die sich mit der Ultraschalldidaktik bei Medizinstudierenden befassten [13,17,48].

Da sich alle Proband:innen in ihrer Konstitution stark ähnelten, fällt auch die Limitation, dass die Prüfungen nicht alle an demselben Schallmodell durchgeführt wurden, nur wenig

ins Gewicht. Ein und dasselbe Schallmodell für alle Prüfungen zur Verfügung zu stellen war aus logistischen Gründen nicht möglich.

#### **4.2.2.5 Videodokumentation**

Ein Vorteil der Videodokumentation gegenüber einer Bewertung in Echtzeit ist die Möglichkeit, die Prüfer:innen zur mehrmaligen Durchsicht aufzufordern. Es ist jedoch zu hinterfragen, wie praktikabel die Bewertung anhand einer Videodokumentation in der Prüfungsrealität an einer Medizinischen Universität ist. Wie eine konkrete Implementierung in ein universitäres Curriculum aussehen könnte, müsste Gegenstand weiterer Untersuchungen sein. Eine interessante Fragestellung für zukünftige Arbeiten wäre zudem, die vergebenen Punkte bei einer Videodokumentation mit denen bei einer Echtzeit-Beurteilung beispielsweise im Rahmen eines Cross-over-Designs zu vergleichen.

#### **4.2.2.6 Bestehensgrenze**

Die Aussagekraft der Anzahl an bestandenen Proband:innen ist limitiert, da die Bestehensgrenze nach üblichen akademischen Standards bei 60% gesetzt wurde und nicht aus bestehenden Daten berechnet wurde. Vor der Implementierung der vorgestellten Prüfung in einen offiziellen curricularen Rahmen sollte eine Untersuchung durchgeführt werden, welche eine konkrete Bestehensgrenze beispielsweise durch die Contrasting-groups-Methode berechnet [40].

#### **4.2.3 Übertragbarkeit der Ergebnisse**

Es wurde lediglich der kurzzeitige Lernerfolg der Proband:innen untersucht. Da die Prüfung direkt nach der dritten und letzten Kurseinheit stattfand, konnten die langfristig beibehaltenen Fähigkeiten nicht überprüft werden. Eine erneute Prüfung im Intervall wäre eine interessante Möglichkeit, sich der Frage der Nachhaltigkeit der erworbenen Fähigkeiten zu nähern. Bei solch einem Studienaufbau gilt es Maßnahmen zu setzen, um die Drop-out-Rate möglichst gering zu halten.

Weiters kann von der kontrollierten Situation in dieser Studie nicht auf die Fähigkeiten der Proband:innen in einem klinischen Setting geschlossen werden. Wie bereits an mehreren Stellen erwähnt war jedoch das Ziel dieser Studie zu evaluieren, in welchem Ausmaß Studierende fähig sind, grundlegende Fertigkeiten der Abdomensonographie im Selbststudium zu erlernen. Es war nicht die Intention, sie auf ein Niveau zu heben, auf dem

sie selbstständig klinisch tätig werden können. Die Ergebnisse sind somit nur eingeschränkt auf didaktische Interventionen im klinischen Setting bei Ärzt:innen übertragbar [31].

Ein gewisser Selektionsbias kann nicht ausgeschlossen werden, da sich potenziell Studierende mit höherer intrinsischer Motivation angemeldet haben. Außerdem stammten alle Proband:innen von einer einzigen Medizinischen Universität. Die Aussagekraft dieser Studie kann somit nur bedingt auf die Gesamtheit der Medizinstudierenden übertragen werden.

### **4.3 Ausblick**

Ein Faktor, der in dieser Studie nicht berücksichtigt wurde, ist der Vergleich der Kosten der beiden Studienarme. Bisher haben nur sehr wenige Arbeiten zur medizinischen Didaktik den Kostenfaktor bei der Einführung neuer Konzepte berücksichtigt [49]. Bei Nilsson et al. (2017) war ein neu konzipiertes E-Learning deutlich weniger kosteneffektiv als ein traditionelles Lehrbuch [29]. Dies führt vor Augen, dass der Kostenfaktor neuer didaktischer Formate kritisch beleuchtet werden sollte.

Ein weiterer Ausblick auf zukünftige Studien wäre die Miteinbeziehung der „Self-directed-learning-readiness-scale“. Dabei wird in einem Fragebogen die Bereitschaft, sich in Eigeninitiative Lerninhalte anzueignen, erhoben [9]. Es wäre eine interessante Fragestellung, die Punkte aus diesem Fragebogen auf Zusammenhang mit den Gesamtpunkten bei einem Selbststudium-Format zu prüfen.

In der bekanntesten Definition von Selbststudium nach Knowles (1975) ist neben dem selbstständigen Wissenserwerb auch die Selbstüberprüfung vorgesehen [8]. Die Fragestellung nach der Übereinstimmung der von Proband:innen selbst wahrgenommenen Leistung und der durch das OSAUS-Bewertungsschema objektivierten Leistung wurde in der existierenden Literatur bisher noch nicht behandelt.

Für zukünftige Studien wäre es dringend anzuraten, die Proband:innen sowohl vor als auch nach dem Kurs zu prüfen, um den Lerneffekt noch besser vergleichen zu können und so Störfaktoren wie unterschiedliche Vorerfahrung zu berücksichtigen. Dies war aus logistischen und zeitlichen Gründen in dieser Studie leider nicht möglich.

## **4.4 Schlussfolgerung**

Ein traditioneller Ultraschallkurs war in dieser Studie dem Selbststudium in der Abdomensonographie nicht signifikant überlegen. Trotz des kontrollierten Settings konnte das Grundproblem, dass zu wenig adäquat ausgebildetes Personal in Relation zur steigenden Nachfrage nach Ultraschallkursen zur Verfügung steht, dahingehend adressiert werden, dass der Betreuungsschlüssel im Selbststudium deutlich reduziert werden konnte.

Abschließend ist zu sagen, dass das Selbststudium in der Ultraschalldidaktik nach der derzeitigen Literatur, welche vorrangig explorativen Charakter hat, nicht als eigenständige Alternative zu traditionellen Kursformaten empfohlen werden kann [48]. Viel mehr sollte das Selbststudium bestehende curriculare Strukturen entlasten und ergänzen, wie es in anderen Feldern der medizinischen Didaktik im Rahmen des Problem-based-Learning bereits gängige Praxis ist [48].

## Literaturverzeichnis

1. Cantisani V, Dietrich C, Badea R, Dudea S, Prosch H, Cerezo E, u. a. EFSUMB Statement on Medical Student Education in Ultrasound [long version]. *Ultrasound International Open*. 11. März 2016;02(01):E2–7.
2. Schmidt G, Görg C. *Kursbuch Ultraschall: Nach den Richtlinien der DEGUM und der KBV*. 6., aktualisierte u. erw. Aufl. Stuttgart: Thieme; 2015.
3. Hoffmann B, Blaivas M, Abramowicz J, Bachmann M, Badea R, Braden B, u. a. Medical Student Ultrasound Education, a WFUMB Position Paper, Part II. A consensus statement of ultrasound societies. *Med Ultrason*. 11. Mai 2020;22(2):220.
4. Prosch H, Radzina M, Dietrich CF, Nielsen MB, Baumann S, Ewertsen C, u. a. Ultrasound Curricula of Student Education in Europe: Summary of the Experience. *Ultrasound Int Open*. Juni 2020;06(01):E25–33.
5. Medizinische Universität Graz, Medizinische Universität Wien, JKU Linz, Medizinische Universität Innsbruck. *Klinischer Lernzielkatalog Österreichs*. 1. Auflage. Graz: Eigenverlag der Medizinischen Universität Graz; 2020.
6. Dietrich CF, Hoffmann B, Abramowicz J, Badea R, Braden B, Cantisani V, u. a. Medical Student Ultrasound Education: A WFUMB Position Paper, Part I. *Ultrasound in Medicine & Biology*. Februar 2019;45(2):271–81.
7. Tarique U, Tang B, Singh M, Kulasegaram KM, Ailon J. Ultrasound Curricula in Undergraduate Medical Education: A Scoping Review. *J of Ultrasound Medicine*. Januar 2018;37(1):69–82.
8. Knowles MS. *Self-directed learning: a guide for learners and teachers*. Chicago: Association Press; 1975.
9. AlRadini F, Ahmad N, Ejaz Kahloon L, Javaid A, Al Zamil N. Measuring Readiness for Self-Directed Learning in Medical Undergraduates. *AMEP*. Mai 2022;13:449–55.
10. Murad MH, Coto-Yglesias F, Varkey P, Prokop LJ, Murad AL. The effectiveness of self-directed learning in health professions education: a systematic review: Effectiveness of self-directed learning. *Medical Education*. November 2010;44(11):1057–68.
11. Garcia-Casasola G, Sánchez FJG, Luordo D, Zapata DF, Frías MC, Garrido VV, u. a. Basic Abdominal Point-of-Care Ultrasound Training in the Undergraduate: Students as Mentors. *J of Ultrasound Medicine*. November 2016;35(11):2483–9.
12. Signor E, Gerstenberger J, Cotton J, Colbert-Getz J, Lappé K. Integrating a self-directed ultrasound curriculum for the internal medicine clerkship. *Ultrasound J*. 5. März 2024;16(1):19.
13. Syperda VA, Trivedi PN, Melo LC, Freeman ML, Ledermann EJ, Smith TM, u. a. Ultrasonography in preclinical education: a pilot study. *J Am Osteopath Assoc*. Oktober 2008;(108(10)):601–5.

14. Mackay FD, Zhou F, Lewis D, Fraser J, Atkinson PR. Can You Teach Yourself Point-of-care Ultrasound to a Level of Clinical Competency? Evaluation of a Self-directed Simulation-based Training Program. *Cureus*. 17. September 2018;10(9):e3320.
15. Cawthorn TR, Nickel C, O'Reilly M, Kafka H, Tam JW, Jackson LC, u. a. Development and Evaluation of Methodologies for Teaching Focused Cardiac Ultrasound Skills to Medical Students. *Journal of the American Society of Echocardiography*. März 2014;27(3):302–9.
16. Canty D, Barth J, Yang Y, Peters N, Palmer A, Royse A, u. a. Comparison of learning outcomes for teaching focused cardiac ultrasound to physicians: A supervised human model course versus an eLearning guided self-directed simulator course. *Journal of Critical Care*. Februar 2019;49:38–44.
17. Höhne E, Recker F, Schmok E, Brossart P, Raupach T, Schäfer VS. Conception and Feasibility of a Digital Tele-Guided Abdomen, Thorax, and Thyroid Gland Ultrasound Course for Medical Students (TELUS study). *Ultraschall Med*. 5. Juli 2021;a-1528-1418.
18. Höhne E, Recker F, Brossart P, Schäfer VS. Teledidactic Versus Hands-on Teaching of Abdominal, Thoracic, and Thyroid Ultrasound—The TELUS II Study. *J GEN INTERN MED*. 12. April 2024;
19. Fuchs L, Gilad D, Mizrakli Y, Sadeh R, Galante O, Kobal S. Self-learning of point-of-care cardiac ultrasound – Can medical students teach themselves? *Lei B, Herausgeber. PLoS ONE*. 27. September 2018;13(9):e0204087.
20. Nielsen M, Cantisani V, Sidhu P, Badea R, Batko T, Carlsen J, u. a. The Use of Handheld Ultrasound Devices – An EFSUMB Position Paper. *Ultraschall in Med*. Februar 2019;40(01):30–9.
21. Sono4You Graz – Ultraschallausbildung von Studierenden für Studierende [Internet]. [zitiert 1. Juli 2024]. Verfügbar unter: <https://graz.sono4you.at/>
22. Celebi N, Zwirner K, Lischner U, Bauder M, Ditthard K, Schürger S, u. a. Student Tutors Are Able to Teach Basic Sonographic Anatomy Effectively – a Prospective Randomized Controlled Trial. *Ultraschall in Med*. 23. November 2010;33(02):141–5.
23. Tolsgaard MG, Todsén T, Sørensen JL, Ringsted C, Lorentzen T, Ottesen B, u. a. International multispecialty consensus on how to evaluate ultrasound competence: a Delphi consensus survey. *PloS one*. 2013;8(2).
24. Höhne E, Recker F, Dietrich CF, Schäfer VS. Assessment Methods in Medical Ultrasound Education. *Front Med*. 9. Juni 2022;9:871957.
25. Ilgen JS, Ma IWY, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ*. Februar 2015;49(2):161–73.
26. Ma IWY, Desy J, Woo MY, Kirkpatrick AW, Noble VE. Consensus-Based Expert Development of Critical Items for Direct Observation of Point-of-Care Ultrasound Skills. *Journal of Graduate Medical Education*. 1. April 2020;12(2):176–84.

27. Gritzmann N, Mathis G. Sonorichtlinien Abdomen – Österreichische Gesellschaft für Ultraschall in der Medizin [Internet]. [zitiert 2. Juli 2024]. Verfügbar unter: <https://oegum.at/zertifizierte-aerztinnen/>
28. Todsén T, Tolsgaard MG, Olsen BH, Henriksen BM, Hillingsø JG, Konge L, u. a. Reliable and valid assessment of point-of-care ultrasonography. *Annals of surgery*. Februar 2015;261(2).
29. Nilsson P, Todsén T, Subhi Y, Graumann O, Nølsøe C, Tolsgaard MG. Cost-Effectiveness of Mobile App-Guided Training in Extended Focused Assessment with Sonography for Trauma (eFAST): A Randomized Trial. *Ultraschall in Med*. Dezember 2017;38(06):642–7.
30. Noerholk LM, Morcke AM, Kulasegaram K, Nørgaard LN, Harmsen L, Andreasen LA, u. a. Does group size matter during collaborative skills learning? A randomised study. *Medical Education*. Juni 2022;56(6):680–9.
31. Todsén T, Jensen ML, Tolsgaard MG, Olsen BH, Henriksen BM, Hillingsø JG, u. a. Transfer from point-of-care Ultrasonography training to diagnostic performance on patients—a randomized controlled trial. *The American Journal of Surgery*. Januar 2016;211(1):40–5.
32. Hofer M, Kamper L, Sadlo M, Sievers K, Heussen N. Evaluation of an OSCE Assessment Tool for Abdominal Ultrasound Courses. *Ultraschall in Med*. 14. Februar 2011;32(02):184–90.
33. Bahner DP, Adkins EJ, Nagel R, Way D, Werman HA, Royall NA. Brightness Mode Quality Ultrasound Imaging Examination Technique (B-QUIET): Quantifying Quality in Ultrasound Imaging. *J of Ultrasound Medicine*. Dezember 2011;30(12):1649–55.
34. Ziesmann MT, Park J, Unger BJ, Kirkpatrick AW, Vergis A, Logsetty S, u. a. Validation of the quality of ultrasound imaging and competence (QUICK) score as an objective assessment tool for the FAST examination. *Journal of Trauma and Acute Care Surgery*. Mai 2015;78(5):1008–13.
35. Black H, Sheppard G, Metcalfe B, Stone-McLean J, McCarthy H, Dubrowski A. Expert Facilitated Development of an Objective Assessment Tool for Point-of-Care Ultrasound Performance in Undergraduate Medical Education. *Cureus*. 10. Juni 2016;8(6):e636.
36. Tiefenböck F. Selbststudium in der Abdomensonographie mittels E-Learning im Vergleich zum Hands-on-Training [Internet]. 2023 [zitiert 12. Mai 2024]. Verfügbar unter: <https://diplomarbeit-abdomensonographie.jimdosite.com/>
37. Ehart G, Hamerschmid A, Leimer E, Queisser L. Abdomen-Sonographie - E-Learning [Internet]. [zitiert 12. Mai 2024]. Verfügbar unter: <https://www.sono4you.at/abdomen/#/>
38. Bundesministerium für Bildung, Wissenschaft und Forschung. *Ars Docendi Preisträger/innen ab 2013* [Internet]. [zitiert 12. Mai 2024]. Verfügbar unter: <https://gutelehre.at/preistraeger/innen-ab-2013>

39. Friedrich Flick Förderungsstiftung. Pocket Ultraschallgeräte für die Ausbildung der kommenden Ärzt\*innengeneration [Internet]. Friedrich Flick Förderungsstiftung. [zitiert 12. Mai 2024]. Verfügbar unter: <https://www.flickfoerderungsstiftung.de/projects/pocket-ultraschallgeraete-fuer-die-ausbildung-der-kommenden-aerztinnengeneration/>
40. Tolsgaard MG, Ringsted C, Dreisler E, Klemmensen, Loft A, Sorensen JL, u. a. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*. April 2014;43(4).
41. Dietrich CF, Fraser AG, Dong Y, Guth S, Hari R, Hoffmann B, u. a. Managing Incidental Findings Reported by Medical, Sonography and Other Students Performing Educational Ultrasound Examinations. *Ultrasound in Medicine & Biology*. Februar 2022;48(2):180–7.
42. Methodenberatung UZH. Datenanalyse mit SPSS [Internet]. [zitiert 13. Mai 2024]. Verfügbar unter: [https://www.methodenberatung.uzh.ch/de/datenanalyse\\_spss.html](https://www.methodenberatung.uzh.ch/de/datenanalyse_spss.html)
43. Cohen J. *Statistical power analysis for the behavioral sciences*. 2. Lawrence Erlbaum Associates; 1988.
44. Krumm S, Schmidt-Atzert L, Amelang M. *Psychologische Diagnostik*. 6., vollständig überarbeitete Auflage. Berlin: Springer; 2021.
45. Dewar ZE, Wu J, Hughes H, Adnani A, Christiansen G, Ovedovitz L, u. a. A comparison of handheld ultrasound versus traditional ultrasound for acquisition of RUSH views in healthy volunteers. *JACEP Open*. Dezember 2020;1(6):1320–5.
46. ÖGUM. Student\*innen-Tutor\*innen Ausbildung [Internet]. [zitiert 1. Juni 2024]. Verfügbar unter: <https://oegum.at/zertifizierungen/#studenten>
47. Bahner DP, Adkins EJ, Nagel R, Way D, Werman HA, Royall NA. Brightness Mode Quality Ultrasound Imaging Examination Technique (B-QUIET): Quantifying Quality in Ultrasound Imaging. *J of Ultrasound Medicine*. Dezember 2011;30(12):1649–55.
48. Boulger C, Prats M, Niku A, Diaz M, Bahner DP. ITSUS: Integrated, Tiered, Self-Directed Ultrasound Scanning for Learning Anatomy. *Cureus*. 2. Juli 2021;13(7):e16119.
49. Zendejas B, Wang AT, Brydges R, Hamstra SJ, Cook DA. Cost: The missing outcome in simulation-based medical education research: A systematic review. *Surgery*. Februar 2013;153(2):160–76.